APPENDIX A

A Gibbs-based unsupervised segmentation approach
to partitioning hyperspectral imagery for terrain applications

Robert S. Rand[*a] and Daniel M. Keenan[*b]

[a] U.S. Army ERDC/ Topographic Engineering Center, Alexandria, VA 22315-3864
[b] Department of Statistics, University of Virginia, Charlottesville, VA 22904

ABSTRACT

A Gibbs-based approach to partitioning hyperspectral imagery into homogeneous regions is investigated for terrain mapping applications. The form of Bayesian estimation, Maximum A Posteriori (MAP) estimation, is applied through the use of a Gibbs distribution defined over a neighborhood system and is implemented as a multi-grid process. Appropriate energy functions and neighborhood graph structures are investigated, which model spectral disparities in an image using spectral angle and/or Euclidean distance. Experiments are conducted on a HYDICE scene collected over an area adjacent to Fort Hood, Texas, that contains a diverse range of terrain features and that is supported with ground truth. Suitable parameter ranges are investigated, and the behavior of the algorithm is characterized using individual and combined measures of disparity within the context of a more general framework, one that supports mixed-pixel processing.

Keywords: Gibbs-based algorithm, Markov Random Field, partitioning, clustering, spectral mixture analysis, hyperspectral imagery.

1. INTRODUCTION

In prior work,[1] an enhanced method of spectral mixture analysis (SMA) was proposed that is appropriate for terrain mapping applications, particularly when the imagery involved is characterized by moderate-to-high scene complexity. This type of complexity is defined to exist whenever many fundamental materials, called "endmembers," exist throughout a scene, or whenever some of the fundamental materials in a scene have spectra that are similar to each other. In such cases, the use of one large set of endmembers for performing SMA can be problematic because of the corresponding high dimensionality of the SMA model's constrained coordinate system, often called its "simplex." A high dimensional simplex can lead to unreliable endmember selection and imprecise estimates of endmember abundance. In order to improve the reliability of the model, we proposed to partition a scene into regions that can be treated as homogeneous for the purposes of spectral unmixing. Once a scene is partitioned into such regions, then we can think of replacing one large set of endmembers by a number of sets containing fewer, and more spectrally distinct, endmembers. Each smaller set of endmembers defines a simplex that is valid for the corresponding partitioned regions, and has a lower dimension than the simplex that accounts for all the endmembers in a scene.

There are many candidate methods for segmenting hyperspectral imagery, both supervised and unsupervised, that could be used to partition a scene into regions. However, one desirable attribute, in the proposed setting, is for the regions to exhibit a certain amount of spatial consistency. Our approach uses a Bayesian framework for image analysis that provides a model for data with two spatial coordinates and one spectral coordinate (i.e., an image cube). A general mathematical development of the underlying theory, applying the Bayesian framework in the context of image processing, is documented.[2] Using such a framework, the properties of a Markov Random Field (MRF) have been applied to develop a Gibbs-based partitioning algorithm for gray-scale imagery with two spatial coordinates, and the Gibbs sampler and Metropolis methods have been used as ways to compute maximum a posteriori (MAP) estimates.[3,4] This development was based on Geman and Geman's pioneering work in applying the MRF framework to the restoration of gray-scale imagery.[4]

The Gibbs-based partitioning approach that we investigate here processes hyperspectral imagery in a way that imposes spatial consistency on the spectral content of sites in each partition. In our previous work,[1] the Gibbs-based partitioning algorithm was designed with an energy function based on a spectral angle disparity metric. This algorithm was shown to be sufficiently sensitive to partition hyperspectral imagery into spatially-consistent regions that are useful for supporting the enhanced

method of SMA, mentioned above, without the clutter often produced by traditional clustering algorithms, such as KMEANS and ISODATA.[5,6] However, the algorithm was not able to distinguish between certain cultural features (asphalt parking lots and certain asphalt-shingled roofs). The sensitivity could be increased to distinguish between these features by lowering the value of a threshold parameter; however, lowering this value would have the undesirable consequence of breaking up natural vegetation features into unwanted partitions.

Herein, two disparity metrics will be investigated – spectral angle and Euclidean distance. A metric shall be defined for each of these measures, and an energy function shall be proposed that can respond to the individual measures, as well as respond to the combined measures (described in Sections 2.2 and 2.3). We seek to increase the sensitivity of the algorithm without introducing undesirable clutter. As an example, we would like the algorithm to distinguish between cultural features without being overly sensitive to uninteresting vegetation changes.

We investigate the Gibbs partitioning algorithm implemented as a multi-grid process, where the processing typically proceeds in stages from course-to-fine grid resolutions and where the beginning (course) grid is initialized randomly. This type of initialization is, in some sense, a worst-case scenario. So the algorithm is totally unsupervised. In practice, one could initialize the algorithm with the some type of preprocessing scheme – perhaps even another unsupervised or supervised classification method.

2. SPECTRAL/SPATIAL PARTITIONING ALGORITHM

2.1 Framework for the Partition Algorithm

A detailed mathematical description of the overall framework for the method discussed here has been formulated,[1] where the partitioning of hyperspectral imagery is used as a mechanism to condition a spectral mixing process. In the current effort, we further investigate the algorithm for estimating the process X^P, called the "partition process," that identifies regions that shall be treated as "homogeneous," in the sense that each region contains some culturally similar phenomenon (not necessarily a single material type).

X^P is a discrete labeling Markov Random Field (MRF) process which associates a label with each site. As in the prior work, we investigate a multi-grid approach where σ is the "grid resolution," which determines the spatial sampling of the algorithm at some specific stage of the multi-grid processing. The term "site" refers to a generic element of a lattice. At full-resolution a site corresponds to a pixel, but at a coarser scale it corresponds to a square block of pixels. The value of σ is defined to be the number of pixels per side of each block of pixels.

We denote S_I as the set of sites on an image lattice at full resolution and $S_P^{(\sigma)}$ as the set of sites on a label lattice at the grid resolution σ:

$$S_P^{(\sigma)} = \left\{ (i\sigma + 1, j\sigma + 1) : 0 \le i \le \frac{N_{Rows} - 1}{\sigma}, 0 \le j \le \frac{N_{Cols} - 1}{\sigma} \right\} \qquad (1)$$

At full resolution $S_P^{(0)} = S_I$. The label lattice $S_P^{(\sigma)}$ has associated graph structures which describe, for a given element, its neighboring elements. These specified structures allow us to sequentially formulate well-defined probability models for X^P as a Markov Random Field (MRF). If S is one of S_I or $S_P^{(\sigma)}$: $S = \{s_1, ..., s_M\}$, then an associated neighborhood system $\mathcal{G} = \{\mathcal{G}_s, s \in S\}$ is specified which is a set of lattice graph structures of neighbors, such that $s \notin \mathcal{G}_s$, and $s \in \mathcal{G}_r$ if and only if $r \in \mathcal{G}_s$.

The structure of the neighborhood system $\mathcal{G}_s$ forms a rather simple pattern. A neighborhood of a site will consist of near, intermediate, and far neighbors, and they are all at specific multiples of σ. Our local neighborhood is either the four closest

(above, below, left, and right), or eight closest (the surrounding perimeter) sites enclosing the site s. The intermediate and far neighbors form a similar pattern as the near neighbors but are at a further distance that is a specific multiple of σ. The intermediate and far neighbors are used to aid in obtaining a faster global solution and to enable the algorithm to remember label assignments computed by the algorithm at prior stages.

We assume a common state space where the realizations of X_s^P are $x_s^P \in \Gamma$, where $\Gamma = \{1, 2, ..., N_{Labels}\}$, and N_{Labels} is the total number of labels assigned to the partitions. The configuration space is then denoted by:

$$\Omega_P^{(\sigma)} = \left\{ x^P = (x_s^P, s \in S_P^{(\sigma)}) \mid x_s^P \in \{1, ..., N_{Labels}\} \right\} \tag{2}$$

X^P is an MRF with respect to the graph $\left\{ S_P^{(\sigma)}, \mathcal{G}^{P(\sigma)} \right\}$, where $\mathcal{G}^{P(\sigma)}$ is the entire neighborhood structure for the partition process at resolution σ, and $\mathcal{G}_s^{P(\sigma)}$ is the specific neighborhood at site s for the partition process at resolution σ. Within this setting, our goal is reduced to finding the value of X^P which maximizes $\Pr(X^P \mid G)$, for the observed G, where G is the hyperspectral image cube.

We use the Gibbs equivalence theorem,[4] and with an appropriately defined graph $\left\{ S_P^{(\sigma)}, \mathcal{G}^{P(\sigma)} \right\}$, we can develop a Gibbs-based algorithm. The advantage of the Gibbs form is that we can compute the MAP estimate, an estimate that maximizes the probability $\Pr(x^P \mid g)$ by iteratively sampling from the local Gibbs distribution pertaining to each site $s \in S_P^{(\sigma)}$. This local distribution is defined as

$$\Pr(X_s^P = x_s^P \mid X_r^P = x_r^P, r \neq s) = \Pr(X_s^P = x_s^P \mid X_r^P = x_r^P, r \in \mathcal{G}_s^{P(\sigma)})$$

$$= \frac{1}{Z_s} e^{-\frac{1}{T} U_s(x_s^P, \underline{g})} \tag{3}$$

where $Z_s = \sum_{x_s^P \in \Gamma} e^{-\frac{1}{T} U_s(x_s^P, \underline{g})}$ and $U_s(x_s^P, \underline{g})$ is the energy interaction of site $s \in S_P^{(\sigma)}$ with the neighborhood $\mathcal{G}_s^{P(\sigma)}$. We use the symbol $\underline{g}$, with the underscore, to emphasize the spatial and spectral nature of the realizations of G, which is in contrast to the spatial-only realizations of the simpler G.[4] Fortunately, because the approach we use to compute the MAP estimate only requires knowledge of the ratios of (3) for current and proposed configurations, the burdensome computation of Z_s does not have to be performed.

We now discuss the specifics of the implementation, below, in Sections 2.2 to 2.4. We define a set of disparity metrics, a number of specific energy functions of the local Gibbs distribution defined by (3), and then discuss a method of computing the corresponding MAP estimates.

2.2 Disparity Metrics

The energy functions that will be introduced in Section 2.3 depend on distance measures, called "disparity metrics," which indicate the dissimilarity between a site s and a neighbor r. We shall investigate disparity metrics that measure distance in terms of the spectral angle metric and the Euclidean distance.

For a pair of sites at the site locations r and s, consider a set $D_{r,s} = \{D_{r,s}^{(1)}, D_{r,s}^{(2)}\}$, where the elements of $D_{r,s}$ are the spectral angle $D_{r,s}^{(1)}$ and the Euclidean distance $D_{r,s}^{(2)}$, between the sites s and r. The individual distance measures $D_{r,s}^{(i)}$ are defined as follows:

$$D_{r,s}^{(1)} = \cos^{-1}\left(\frac{\underline{g}_s \cdot \underline{g}_r}{|\underline{g}_s| \cdot |\underline{g}_r|}\right) \tag{4}$$

$$D_{r,s}^{(2)} = sqrt\left(\frac{(\underline{g}_s - \underline{g}_r)^t \cdot (\underline{g}_s - \underline{g}_r)}{N_{bands}}\right) \tag{5}$$

2.3 Energy Functions

Energy functions shall now be defined that respond to disparities as measured by the spectral angle metric, the Euclidean distance and perhaps others. As mentioned earlier, we seek to increase the sensitivity of the algorithm without introducing undesirable clutter.

For each individual disparity metric in (4) – (5), an energy function is defined as

$$U_s^{(i)}(x_s^P, \underline{g}) = \sum_{r \in \mathcal{G}_s} \delta(x_s^P, x_r^P) \cdot f_i(D_{r,s}^{(i)}) \qquad i=1,2 \tag{6}$$

where the delta function $\delta(x_r, x_s) = 1$ for $x_r = x_s$, and $\delta(x_r, x_s) = 0$ for $x_r \neq x_s$.

The function $f_i(D_{r,s}^{(i)})$ maps $D_{r,s}^{(i)}$ to the set $\{-1, 1\}$, where $f_i(D_{r,s}^{(i)}) = -1$ indicates a non-significant distance between r and s, and a value $f_i(D_{r,s}^{(i)}) = 1$ indicates a significant distance. If $D_{r,s}^{(i)} \leq K_{thresh}^{(i)}$, then $f\left(D_{r,s}^{(i)}\right) = -1$, and otherwise $f\left(D_{r,s}^{(i)}\right) = 1$, where the threshold $K_{thresh}^{(i)}$ determines the sensitivity of the algorithm to the $D_{r,s}^{(i)}$ distance measure. The effect of this energy function is to accumulate the value of $f\left(D_{r,s}^{(i)}\right)$ whenever the label of a site s is equal to its neighbor at r.

For $i=1,2$, the value of the threshold $K_{thresh}^{(i)}$ is determined through observation; specifically, $K_{thresh}^{(1)}$ is determined by observing the differences in spectral angle for samples of the same material verses the differences in spectral angle for samples of different materials, and $K_{thresh}^{(2)}$ is determined by observing the differences in Euclidean distance for samples of the same material verses the differences in Euclidean distance for samples of different materials. If the imagery is calibrated to reflectance, then the value attained should be fairly universal, presuming a representative set of materials has been analyzed. If the imagery has not been calibrated, then this value will depend on the range and scaling of the uncalibrated data.

In addition to the energy functions (6) that respond to a single disparity metric, we can construct an energy function that responds to a pair of disparity metrics:

$$U_s(x_s^P, \underline{g}) = \sum_{r \in \mathcal{G}_s} \delta(x_s^P, x_r^P) \cdot \max\{f_1(D_{r,s}^{(1)}), f_2(D_{r,s}^{(2)})\} \tag{7}$$

2.4 Calculating the MAP estimate

A simulated annealing technique is used to obtain the MAP estimate via the Metropolis algorithm,[1,2,3,4,8] with an eventual fast freeze at low temperature. The computations involve a temperature t_k at iteration k that defines an annealing schedule. We use a commonly applied annealing schedule,[2,4] namely,

$$t_k = \frac{C}{\ln(1 + k)} \tag{8}$$

C is a parameter that corresponds to the maximum depth of a "well" in the energy function (6) or (7).[2] The setting for C affects the rate of convergence. For example, an upper bound on this value for some disparity i can be computed as the value of (6), for which the maximum distance $f_i(D_{r,s}^{(i)})$, of the site-neighbor pairs are all different, but the labels are assigned the same value. When running the multi-grid algorithm, computing time can be reduced considerably by using a large C value at the coarsest grid, and otherwise using smaller C values at the finer grids. Practical values of C are discussed in Sections 3 and 4.

3. DESCRIPTION OF EXPERIMENT

3.1. Data

A HYDICE scene is used in the experiment that was extracted from data collected over a geographic area near Fort Hood, Texas at a 10,000 feet altitude, on 16 May 1996. The scene consists of 300 samples by 600 lines of pixel vector data with a spatial resolution of approximately 3 meters. Complete HYDICE image cubes have 210 bands in the approximate 0.4 –2.5 μm spectral range. However, because of the severe degradations that occur in atmospheric windows of the spectrum and sensor degradations occurring in certain bands, a maximum of 117 bands were used to compute the disparities within the energy functions. These 117 bands of data were calibrated to reflectance by the Empirical Line Method based on the known spectra of the calibration panels that were located in a region outside this particular scene. Ground truth, describing certain ground features in the scene, is based on data collected during a site visit to the area during April 2000. Figure 1 shows a ground-truth overlay superimposed over Band 49 extracted from the HYDICE image cube. Descriptive information pertaining to the ground truth is listed in Table 1.
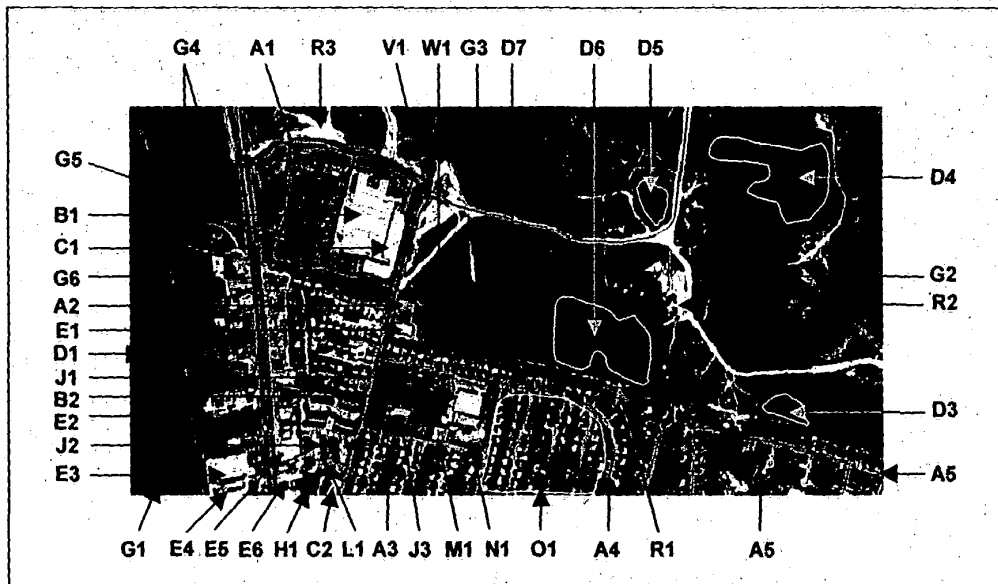

Figure 1. HYDICE Scene (Band 49) with Ground-Truth Overlay

Because of the time difference between scene acquisition and the site visit, not all portions of the scene could be documented with certainty. In addition to the ground-feature data, spectral data of a diverse range of materials, measured by field spectrometers and collected independently of this experiment at other geographic locations, were used to determine an appropriate quantization threshold for the spectral angle and Euclidean distance disparity metrics, as discussed below. Ground-measured spectral data within the study site were not available.

3.2. Validation of the Algorithm

An experiment was conducted using computer code written in C++ that implements our algorithm in the manner described in Section 2. The energy functions in (6) and (7) were implemented with the disparity metrics defined in (4)-(5). In order to determine appropriate values of the thresholds $K_{thresh}^{(1)}$ and $K_{thresh}^{(2)}$, the spectral angles and Euclidean distance measures between all the pairwise combinations of the spectra in the reference spectral library, just mentioned above, were computed.

Table 1. Description of the ground truth (GT) information, as displayed in Figure 1.

GT Label	Description of the Area
A1	Asphalt parking lot, newly paved.
A2-A3	Asphalt parking lot. A3a and A3b opposite sides of apartment complex, different ages.
A4-A5	Asphalt road
B1	Rooftop of a department store, light-gray asphalt shingles, skylights, yellow gas pipes.
B2a,b	School, rooftop mostly coated bubbly light-gray material (B2b), top edge is gravel (B2a).
C1	Concrete pavement to loading dock, some tire marks.
C2a,b	Concrete parking lot surrounding small bank building. C2b is brighter than C2a.
D1	Trees: Black Willow, Texas Sugar Berry, Dogwood, Texas oak, Elbon bush, Green Bryer.
D3	Trees
D4	Trees: Texas Oak, Burmelia, and Texas Hackberry.
D5	Trees: Red Bud, Texas Oak, and Red Bud.
D6a,b,c	Trees - Deciduous holly, Bois D'Arc, Osage Orange, Texas Oak, and Bow Wood
D7	Scattered junipers throughout the grassy areas.
E1	Church and school with metal rooftop.
E2-E6	Corrugated steel buildings.
G1-G6	Tall wild grass. G5 and G6 near drainage and healthier.
H1	Healthy maintained grass near bank building.
J1	Private residence, stand-alone.
J2	Apartment buildings with asphalt-shingled rooftops.
J3	Church with dark brown asphalt-shingled rooftop.
L1	Shade of Apartment building.
M1	Tennis court, concrete with blue/gray coating, and lots of dirt.
N1	Playground area with exposed soil.
O1	Encircled area is residential with moderately-priced homes and concrete driveways.
R1	Light linear pattern is exposed bedrock.
R2-R3	Exposed soil.
V1	Asphalt road intersection with exposed soil along the shoulders.
W1	Playground - grass with exposed soil, gravel, and some asphalt.

The experiment consisted of two trials. Trial A used the spectral angle disparity metric in the energy function and Trial B used a combined spectral angle and Euclidean distance metric in the energy function. The behavior of the algorithm is affected by various parameters; namely, σ (the grid resolution), C (controls the annealing schedule through the temperature t_k), K_{total} (final number of iterations through the image), and N_{Labels} (the number of partition labels). Table 2 lists the values of these parameters. As indicated in Table 2, the trials proceeded in a multi-grid sequence with grid resolutions

$\sigma = 16, 8, 4, 2, 1$. The unsupervised ISODATA clustering method is used as a reference method. Figure 2 shows the results of ISODATA on the data for 6 classes.

Table 2. Control parameters for Trials A and B. The sequence of numbers in the "Dist of neighbors" column corresponds to the distance from the center pixel (the "site") to the perimeter, for the surrounding near neighbors, the intermediate neighbors, and the far neighbors. The sequence of numbers in the "# of neighbors" column corresponds to the number of near neighbors, intermediate neighbors, and far neighbors in the neighborhood system.

σ	C	# iterations	Dist. of neighbors	# of neighbors
16	9.0	50,000	16,64,128	8, 8, 8
8	3.0	3,000	8,32,64	8, 8, 8
4	3.0	3,000	4,16,32	8, 8, 8
2	2.0	700	2,8,16	8, 8, 8
1	1.0	350	1,4, 8	8, 8,8

During the trials, a range $C = (1.0\text{-}9.0)$ was investigated, which is below the upper bound discussed in Section 2. In general, the value of C was set to its highest value at the initial stage of a run (corresponding to high temperature) and it was set to lower values for the subsequent stages. In addition to the practical computational advantage, a lower C value allows the algorithm to "remember" the results at previous stages. The distances of the intermediate and far neighbors shown in Table 2 corresponds to fixed multiples of the grid resolution, 4σ and 8σ, respectively. For all the trials, the algorithm was initialized randomly, and then proceeded from a grid resolution of $\sigma=16$ to $\sigma=1$.

4. RESULTS

Figure 2 and Table 3a shows the results of the ISODATA clustering method applied to the image data for 6 classes. Although ISODATA provides a fairly good partitioning of the scene, it is overly sensitive to minor vegetative differences in the forested areas, denoted by "D1-D6" in the ground-truth map (Figure 1) as well as by shade. In particular, Table 3a shows a very important problem, where another label associated with A1 (asphalt) is prevalent throughout subregions of D1-D6 that can be attributed to shade. Specifically, 27% (1015 out of 3764) of the D1-D6 pixels are labeled with the same cluster associated with asphalt. Also 15.7% (559 out of 3559) of the G1-G4 pixels are labeled with the same cluster as associated with trees. Furthermore, the clustering does not provide a smooth and spatially consistent mapping of regions, and exhibits a significant amount of speckling.

Based on observations of spectral angle distances between the various pairs of materials of our previous work,[1] we set $K^{(1)}_{thresh} = 11.0$, which remained constant throughout the partition trials. Based on observations of Euclidean distances between the various pairs of materials, we set $K^{(2)}_{thresh} = 100.0$, which also remained constant throughout the partition trials.

For both metrics and most of the pair-wise signature combinations, there is an obvious gap between similar materials and different materials, except for combinations involving a vegetation signature. Unfortunately, vegetation has a fairly large variance, which makes the threshold boundary a bit fuzzy. During these trials, we observed that the exact value of these thresholds did not seem to be very critical.

Figure 3 shows the results of Trial A testing the spectral angle metric in the energy function. On the positive side, the algorithm with this energy function was successful in characterizing the structure of many of the natural and cultural features in the scene without much of the undesirable clutter prevalent in the ISODATA scene shown in Figure 2. For example, the algorithm was not overly sensitive to shade differences such as in the forested areas, denoted by "D1-D6" in the ground-truth map. Unfortunately, the algorithm was also insensitive to other differences that we would like to discriminate. For example, there is missing structure within the road intersection denoted by "V1," and the rooftops of the apartment buildings (asphalt shingles) denoted by "J2" are confused with the surrounding asphalt parking lot "A3."

Table 3. Confusion matrices of results. Labels in the 18 rows are the GT labels, and labels in the 6 columns are the partitioned labels. As an example of how to read the tables: In Table 3a, the region of ground truth extracted for C1, consisted of 120 pixels. Of the 120 pixels, 13 pixels were assigned to ISODATA cluster label ISO-04, and 107 pixels were assigned to ISODATA cluster label ISO-05.

Table 3a. Results for the ISODATA algorithm.

Labels	ISO-01	ISO-02	ISO-03	ISO-04	ISO-05	ISO-06
A1	0	0	518	0	0	0
B1	0	0	0	278	0	0
C1	0	0	0	13	107	0
C2a	0	0	0	0	13	0
C2b	0	0	0	0	11	0
B2a	0	0	0	2	15	0
B2b	0	0	0	12	110	0
M1	0	0	0	43	0	0
J3	0	0	33	44	0	0
A2	0	4	0	51	0	0
A3a	0	0	0	16	0	0
A3b	0	0	0	12	0	0
J2	0	0	84	0	0	0
D1-6	2276	36	1015	0	0	1
G1-G4	559	2989	3	0	0	8
H1	0	0	0	0	0	33
G5	101	0	0	0	0	71
G6	0	10	0	0	0	176

Table 3b. Results for the Gibbs-based algorithm using the combined disparity (CD) metrics.

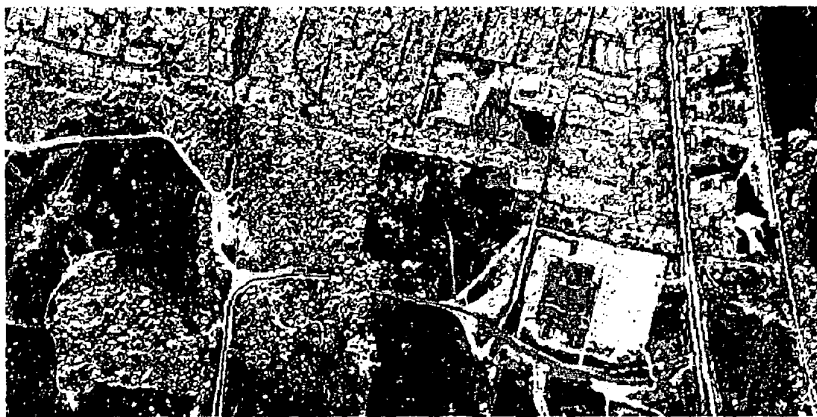| Labels	CD-05	CD-01	CD-02	CD-04	CD-03	CD-06
A1	0	0	518	0	0	0
B1	0	0	0	278	0	0
C1	0	0	0	0	120	0
C2a	0	0	0	7	0	0
C2b	0	0	0	11	0	0
B2a	0	0	0	17	0	0
B2b	0	0	0	122	0	0
M1	41	0	0	0	2	0
J3	0	0	77	0	0	0
A2	0	0	55	0	0	0
A3a	0	0	0	0	16	0
A3b	0	0	0	0	12	0
J2	0	0	84	0	0	0
D1-6	3527	66	26	62	48	35
G1-G4	1	3543	0	1	0	14
H1	33	0	0	0	0	0
G5	0	0	0	0	0	172
G6	0	185	0	0	1	0

Figure 4 and Table 3b shows the results of Trial B using both the spectral angle and Euclidean distance metrics in the energy function. A comparison of Figures 3 and 4 shows a clear advantage to using the two metrics in the energy function. Many of the cultural structures missed in Trial A are detected in this trial without introducing undesirable vegetation clutter. For example, the structure in "V1" is now evident and so are the rooftops in "J2." In addition, Table 3b shows that the regions D1-D6 are labeled quite homogeneously and are not adversely affected by shade, as was the case with ISODATA. Specifically, 0.7% (26 out of 3764) of the D1-D6 pixels are labeled with the same cluster associated with asphalt. Also 0.4% (16 out of 3559) of the G1-G4 pixels are labeled with the some other cluster (trees or otherwise).

The above trials corresponding to Figures 3-4 were performed using identical control parameters, as listed in Table 2, in order to provide some consistency for comparing the different methods. The results indicate that an energy function using the combined spectral angle and Euclidean distance metrics is the preferred method.

Overall, the Gibbs-based algorithm generates partitions that accurately represent structured regions in the image. For most of the bigger regions, pixels with the same label represent the same type of phenomenon on the ground, globally across the image, and regions with different labels correspond to different phenomenon. However, a true globally consistent labeling of the partitions was sometimes problematic. In particular, labeling problems can be seen in some of the smaller regions. Although the contiguous pixels in smaller partitions represent the same phenomenon, locally, some small partitions, separated by a certain amount of distance and given the same names, actually represent different phenomenon. Conversely, some small partitions, separated by a certain amount of distance and given different names, actually represent similar phenomenon.

5. CONCLUSIONS

Two energy functions were tested involving the use of the spectral angle metric, and a combined spectral angle metric and Euclidean distance, respectively. The results of our experiment indicate that the energy function using a combination of the spectral angle and Euclidean distance metrics produced the best results.

Overall, the Gibbs-based algorithm generates partitions that accurately represent structured regions in the image. A globally consistent labeling of the partitions was sometimes problematic, particularly, for some of the smaller regions. This global labeling problem can be significantly reduced by using a non-random initialization scheme, or perhaps by performing more than one pass through the multi-grid sequence.

The Gibbs-based algorithm is envisioned to work within an overall framework involving other processes. In particular, the partitioning of hyperspectral imagery is viewed as a mechanism to condition a spectral mixing process. Under this framework, the ultimate identification of materials (and the corresponding abundances) in a scene is a consequence of the spectral mixing process, so that a precise global labeling is not required at the partitioning stage.

ACKNOWLEDGEMENTS

REFERENCES

[1] R. S. Rand and D. M. Keenan, "A Spectral Mixture Process Conditioned by Gibbs-based Partitioning," *IEEE Transactions on Geoscience and Remote Sensing: Special Issue on the Analysis of Hyperspectral Image Data*, submitted for publication, 2001.

[2] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods – A Mathematical Introduction,* Applications of Mathematics, Springer-Verlag, New York, 1995.

[3] D. Geman, S. Geman, C. Graffigne., and P. Dong, "Boundary detection by constrained optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 7, July 1990.

[4] D. Geman and S. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, November 1984.

[5] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 1973.

[6] C.W. Therrien, *Decision Esimation and Classification – An Introduction to Pattern Recognition and Related Topics*, John Wiley and Sons, New York, 1989.

[7] R. V. Hogg and E. A. Tannis, *Probability and Statistical Analysis*, 3rd Ed., Macmillan Publishing Company, New York, 1988.

[8] E. Aarts and J. Korst, *Simulated Annealing and Bolzmann Machines - A Stochastic Approach to Combinatorial Optimization and Neural Computing*, Interscience Series in Discrete Mathematics and Optimization, John Wiley and Sons, 1989, Reprint 1990.

APPENDIX B

A Spectral Mixture Process Conditioned by Gibbs-Based Partitioning

Robert S. Rand, *Member, IEEE,* and Daniel M. Keenan

Abstract—An enhanced method of spectral mixture analysis is investigated for hyperspectral imagery of moderate-to-high scene complexity, where either a large set of fundamental materials may exist throughout, or where some of the fundamental members have spectra that are similar to each other. For a complex scene, the use of one large set of fundamental materials as the set of "endmembers" for performing spectral unmixing can cause unreliable estimates of material compositions at sites within the scene. In such cases, partitioning this large set of endmembers into a number of smaller sets is appropriate, where the smaller sets are associated with certain regions in a scene. Herein, a Gibbs-based algorithm is developed to partition hyperspectral imagery into regions of similarity. This partitioning algorithm provides an estimator of an underlying and unobserved process called a "partition process" that coexists with other underlying (and unobserved) processes, one of which is called a "spectral mixing process." The algorithm exploits the properties of a Markov random field (MRF) and the associated Gibbs equivalence theorem, using a suitably defined graph structure and a Gibbs distribution to model the partition process. Consequently, spatial consistency is imposed on the spectral content of sites in each partition. The enhanced spectral mixing process is then computed as a linear mixture model that is conditioned on the partition process. Experiments are performed using scenes of HYDICE imagery to validate the algorithm, where spectral mixture analysis is performed with and without conditioning on the partitioning process.

Index Terms—Bayesian, hyperspectral, multigrid Gibbs distribution, partition process, spectral mixture analysis.

I. INTRODUCTION

BEFORE introducing our approach, we first give a broad overview of the diverse array of detection/classification/identification techniques which have typically been applied to hyperspectral imagery. They are inherently either full-pixel or mixed pixel techniques, where each pixel vector in an image records the spectral information. The underlying assumption governing full-pixel techniques is that each pixel vector measures the response of predominantly one underlying material/signal at each site in a scene. In contrast, the underlying assumption governing mixed-pixel techniques is that each pixel vector measures the response of multiple underlying materials/signals at each site in a scene. Unfortunately, an image can often be a combination of the two situations, where many sites in a scene are pure materials, but many others are

Manuscript received November 8, 2000; revised January 30, 2000.
R. S. Rand is with the U.S. Department of the Army Engineer Research and Development Center (ERDC), U.S. Corps of Engineers, Topographic Engineering Center, Alexandria, VA 22135 USA (e-mail: robert.s.rand@erdc.usace.army.mil).
D. M. Keenan is with the Department of Statistics, University of Virginia, Charlottesville, VA 22903 USA (e-mail: dmk76@virginia.edu).
Publisher Item Identifier S 0196-2892(01)05503-6.

mixtures of materials. For perspective, we briefly discuss some of the full-pixel and mixed-pixel techniques, and then introduce the enhanced spectral mixture analysis concept.

By construction, the modeling and algorithms investigated here are intended for mapping applications, rather than targeting applications. The approach encourages spatial integrity of any assigned class labels, so, consequently, the detection of single-pixel and subpixel anomalies are discouraged, unless there is strong spectral evidence to the contrary. Also, of possible interest is the following: although the current effort focuses on the use of a Gibbs-based partitioning algorithm for enhanced spectral mixture analysis, this partitioning algorithm can also be used with any of the full-pixel methods mentioned, to add spatial integrity to an otherwise spectral-only method, as discussed later.

A. Full-Pixel Techniques

The simplest and most straightforward full-pixel technique is the method of spectral matching. Spectra of interest in an image are matched to training spectra that are either obtained from a reference library of spectra or from the image itself. Simple metrics for determining the degree of match are such measures as Euclidean distance, derivative difference, and spectral angle [1], [2]. Despite its simplicity, spectral matching can be very effective as long as the training data are properly calibrated with respect to data of interest and provided that the full-pixel scenario is appropriate. Unfortunately, a certain number of the pixel vectors in a scene will likely measure the spectral response of a mixture of materials. If the number of the mixed pixels is significant with respect to the size of the scene then spectral matching should not be used, at least not as an isolated technique. Also, class label assignments generated by spectral matching algorithms are not affected by spatial neighborhoods, which may be a positive consequence for targeting applications. However, for mapping applications, consistency of class labels in localized spatial neighborhoods, termed as "spatial localization," is important.

Other full-pixel methods that have been applied to hyperspectral imagery are the various supervised and unsupervised scene segmentation techniques, which are based on statistical and pattern recognition methods that were previously used for lower-dimensional multispectral processing. As with spectral matching, the required training is accomplished with data from either spectral libraries or imagery, again with the same caveats. The techniques include statistical linear discrimination (e.g., Fisher's linear discriminant), quadratic multivariate classifiers (e.g., Mahalanobis and Bayesian maximum likelihood [ML] classifiers) [3]–[5], and neural networks [6]. The quadratic methods require low-dimensional pixel vectors, and hence,

these techniques are typically preceded by a data reduction operation to reduce the number of bands. Such neural networks as the multilayer feedforward neural network (MLFN) can be constructed to model quadratic and higher-order decision surfaces without the need to reduce the number of bands, and assessments of such methods have been performed [7]–[10]. Studies comparing neural networks to traditional quadratic classifiers have also been performed on multispectral imagery, with a general consensus favoring the neural methods [11], [12]. An advantage of certain neural techniques such as the MLFN, is that they can be trained to identify materials that have been perturbed by a limited amount of mixing. The methods usually do not include any type of spatial localization in the decision making process.

The most commonly applied unsupervised algorithms for clustering multispectral and hyperspectral imagery are KMEANS and ISODATA [4], [5], where the metric used in determining the cluster membership is the Euclidean distance. Unfortunately, this metric is not particularly responsive to the fine structure, or shapes, in high-resolution spectra, and is often overly sensitive to intensity differences. Also, these two methods do not include any type of spatial localization in the clustering operation.

B. Mixed-Pixel Techniques

Spectral mixture analysis (SMA) techniques have overcome some of the weaknesses of full-pixel approaches, by using linear statistical modeling and/or signal processing techniques [13]–[17]. The basic equation is

$$X_s = H\beta_s + \eta_s \tag{1}$$

in matrix notation, where, at site s in the scene, X_s is the observed reflected energy, β_s is the modeling parameter vector that will be a associated with the mixture proportions, and η_s is the random variable for model error. The columns of matrix H contain the spectra of the pure materials, called "endmembers." A fundamental aspect of basic SMA is that the matrix H is presumed known and fixed (nonrandom). The assumption of a fixed H is problematic, because for most real-world materials, there are no single fixed spectral signatures to populate H that can precisely represent the pure materials. Further aspects of SMA are discussed in Section III-B.

Modifications to the basic SMA method often involve partitioning the matrix H in (1) into desired and undesired signatures and then computing subspace projections that are orthogonal and/or oblique to the undesired signature and noise components. For example in [18], the concept of orthogonal subspace projection (OSP) is applied to hyperspectral imagery that suppresses undesirable signatures and detects a specific signature of interest. In this method, $H = [D, U]$, where D is the known spectra for a target of interest and U is a matrix of undesired (but known signatures). In [19], a modification of this technique is presented to detect the presence of a known signature in the presence of an unknown and undesired background, where the columns of U are unknown. However, D has to be a minor component of the scene [19]. These modifications to the basic SMA

and other similar ones [20] are best suited for targeting applications.

Unless we have accurate ground-truth information about the materials in a scene, the task of determining the number of endmembers in a scene and subsequently identifying these endmembers is not trivial. Some progress has been made in cases where ground truth is not available, but reliable solutions remain allusive, particularly for scenes of moderate-to-high complexity. Usually, the first step is to "noise whiten" the data and then assess its dimensionality. A popular technique is the minimum noise fraction (MNF) transform [21]. The MNF transform requires estimating a noise covariance matrix $\sum_n$, which is typically done using the minimum/maximum autocorrelation factors (MAF) method developed in [22]. A noise adjusted principal component (NAPC) method was later shown to be mathematically equivalent to the MNF but easier to compute [23]. An interactive method using MNF transformed data to identify the endmembers has been developed, based on convex geometry concepts [24], [25] and the minimum volume transform [26]. An automated method has been proposed in [27], that of applying the minimal descriptive length (MDL) criterion to noise whitened data for determining the number of endmembers, followed by an orthogonal subspace projection method for identifying them. More recently, methods of independent component analysis (ICA) have been investigated [28], [29]. Unfortunately, the reliability of these methods hinges on an ability to accurately estimate $\sum_n$. Difficulties in obtaining reliable estimates of $\sum_n$ are discussed within a targeting framework in [30], in the context of comparing MNF and PCA transform methods.

C. Proposed Approach

The present investigation considers hyperspectral image analysis from a broader perspective than the individual methods mentioned above. Our approach will use the paradigm of Bayesian image analysis. A treatment of the various aspects of Bayesian image analysis is presented in [31], such as the Bayesian paradigm, the Gibbs sampler, and methods of estimation. Additional mathematical development of the underlying theory, applying the Bayesian framework to Random Processes in the context of image processing, is presented in [32]. Bayesian approaches using Markov Random Field (MRF) models have provided algorithms for producing unsupervised partition maps and detecting boundaries on single-band imagery [33]. Algorithms by Geman et al. [34] use the Gibbs sampler and Metropolis methods to compute maximum a posteriori (MAP) estimates. Recently, a Gaussian MRF approach that considers the spatial-spectral correlation structure of multispectral imagery using the Iterative Conditional Mode (ICM) has been developed [35]. Although multiband imagery has been treated to some extent in the Bayesian community, the significantly different character of hyperspectral imagery has not really been addressed.

In our experiments, we investigate the Gibbs partitioning algorithm as an unsupervised method, where there is random initialization (in some sense, a worst-case scenario). In practice, one can consider potential preprocessing methods for initialization. One could initialize the proposed algorithm with one of the

full-pixel techniques mentioned above, in which case the algorithm can effectively become supervised. Although this study limits itself to random initialization, we see these alternative initialization schemes as an important topic for future investigation. In this manner, our approach of incorporating spatial localization can potentially improve the results of a number of spectral-only techniques.

An enhanced method of spectral mixture analysis is developed that should be appropriate for scenes of moderate-to-high scene complexity. This type of complexity is defined to exist whenever many endmembers exist throughout a scene, or whenever some of the fundamental members have spectra that are similar to each other. For a complex scene, the use of one large set of endmembers for performing SMA can cause unreliable estimates of material compositions at sites within the scene. Our approach is to partition such a large set of endmembers into a number of sets containing fewer, and more spectrally distinct, endmembers, where the smaller sets are associated with certain regions in a scene. Recalling the above mentioned fundamental (and historical) difficulties in automating endmember selection, during the experiments we will be extracting our endmember spectra manually, and assigning them to sets based on known ground-truth information, in order to control sources of error.

The remainder of the paper is organized as follows. A mathematical description of the Bayesian model is presented, followed by a description of our Gibbs-based algorithm developed to partition hyperspectral imagery in an unsupervised manner. A description of the experiments using HYDICE imagery is given for both validating the algorithm and computing the spectral mixture process conditioned by the partition process.

II. MATHEMATICAL FRAMEWORK

A Bayesian model is introduced in this section to account for spectral mixing and the partitioning of homogeneous regions in hyperspectral imagery. The approach is model-based. Multiple conditional probability structures are used to bind together a hierarchy of underlying and unobserved processes that combine to generate a given scene. We assume that there is an underlying hierarchy of two processes

$$X = (X^F, X^P). \tag{2}$$

The process X^P, called the partition process, will identify regions that shall be treated as homogeneous. The process X^F, called the spectral mixing process (SMP), will model the reflected energy, possibly being that of a composite of materials.

Then, what is actually observed is G, the observed hyperspectral image cube

$$"G = X^F + \text{noise}." \tag{3}$$

For each of G and X^F, there are two "coordinates," one spatial and one spectral.

Notationally, we make the following definitions. Let S_I denote the full-resolution image pixel lattice, i.e., the resolution at which the observations are obtained, which will also denote the full-resolution partition lattice S_P

$$S_P = S_I = \{(i,j): 1 \leq i \leq N_{\text{Rows}}, 1 \leq j \leq N_{\text{Cols}}\} \tag{4}$$

where N_{Rows} is the number of rows in an image, and N_{Cols} is the number of columns. Let Λ denote the full-resolution wavelength lattice

$$\Lambda = \{\lambda_l : \lambda_l \in (400\text{ nm},\ 2500\text{ nm}), l = 1, N_{\text{bands}}\} \tag{5}$$

where N_{bands} is the number of spectral bands in the hyperspectral image cube. Hence, for example, the full-resolution coordinates for the observed hyperspectral G are $S_I \times \Lambda$:$s \in S_I, \lambda \in \Lambda$. Therefore, $G(s, \lambda)$ is the observed intensity at that "site" and wavelength.

Below, we present the constructions of X^P, X^F, and G. Many of the explicit details are given in Section III.

A. Construction of the X^P Process

As stated in Section I, our primary objective is to develop a method for image segmentation that utilizes both spatial and spectral information. Hence, the construction of X^P is fundamental. Our approach for the spatial-spectral partitioning will be multigrid, and thus we define a collection of sublattices of S_P for the partition process

$$S_P^{(\sigma)} = \left\{ (i\sigma + 1, j\sigma + 1) : \\ 0 \leq i \leq \frac{N_{\text{Rows}} - 1}{\sigma}, 0 \leq j \leq \frac{N_{\text{Cols}} - 1}{\sigma} \right\} \tag{6}$$

where σ is a parameter called the "grid resolution," which determines the spatial resolution (coarseness) of the labeling process at some level in the hierarchy.

Below, the term "site" will refer to a generic element of one of the aforementioned lattices. At full-resolution it would be a pixel, but on a cruder scale, it will be a square block of (full-resolution) pixels. The value of σ is defined to be the number of pixels per side of each block of pixels. A site might be composed of one material, but it could possibly be composed of numerous materials, and we need to presume that we will not know this, or the identities of the materials, ahead of time. Moreover, the above sets, S_I and $S_P^{(\sigma)}$, will have associated graph structures which describe, for a given element, its neighboring elements. These specified structures allow us to sequentially formulate well-defined probability models for X^P (as a Markov random field [MRF]) as well as for X^F and G.

Notationally, if S is one of S_I or $S_P^{(\sigma)}$: $S = \{s_1, \ldots, s_M\}$, where M is the number of sites on the grid, then an associated neighborhood system $\mathcal{G} = \{\mathcal{G}_s, s \in S\}$ is specified (details in Section III) which is a set of lattice graph structures of neighbors such that $s \notin \mathcal{G}_s$, and $s \in \mathcal{G}_r$ if and only if $r \in \mathcal{G}_s$. Within this setting, our goal is reduced to finding the value of X^P, which maximizes $\Pr(X^P|G)$ for the observed G.

X^P is a discrete labeling process that associates a label with each site and is treated as an MRF. The realizations of X_s^P are $x_s^P \in \Gamma$, where $\Gamma = \{1, 2, \ldots, N_{\text{Labels}}\}$. We are using the general notation of upper-case "X" being a random variable and lower-case "x" being a realizable value. N_{Labels} is the total number of labels assigned to the partitions. The configuration

space for partitioning is then determined by N_{Labels} and σ (the label resolution), and denoted by

$$\Omega_P^{(\sigma)} = \left\{ x^P = \left(x_s^P, s \in S_P^{(\sigma)} \right) | x_s^P \in \{1, \ldots, N_{\text{Labels}}\} \right\}. \tag{7}$$

$S_P^{(\sigma)}$ is the set of sites used in the partition process at a resolution σ. The development for X^P is similar to that of [33] with the important addition of a spectral dimension.

We treat X^P a Markov random field (MRF) with respect to a graph $\{S_P^{(\sigma)}, \mathcal{G}^{P(\sigma)}\}$ so that

$$\Pr(X^P = x^P) > 0 \text{ for all } x^P \in \Omega_P^{(\sigma)} \tag{8}$$
$$\Pr\left(X_s^P = x_s^P | X_r^P = x_r^P, r \neq s\right)$$
$$= \Pr\left(X_s^P = x_s^P | X_r^P = x_r^P, r \in \mathcal{G}_s^{P(\sigma)}\right) \tag{9}$$

where $\{X^P = x^P\}$ means $\{X_{s_1}^P = x_{s_1}^P, \ldots, X_{s_M}^P = x_{s_M}^P\}$, $\mathcal{G}^{P(\sigma)}$ is the entire neighborhood structure for the partition process at resolution σ, and $\mathcal{G}_s^{P(\sigma)}$ is the specific neighborhood at site s for the partition process at resolution σ.

We exploit the Gibbs equivalence theorem offered in [34] and with an appropriately defined graph $\{S_P^{(\sigma)}, \mathcal{G}^{P(\sigma)}\}$, we can adopt a Gibbs representation

$$\Pr(X^P = x^P, G = g) = \frac{1}{Z} e^{-\frac{1}{T} U(x^P, g)} \tag{10}$$

where $Z = \sum_{x^P} e^{-(1/T)U(x^P, g)}$ is known as the partition function of statistical physics (not to be confused with the "partition process") and $U(x^P, g)$ is the total energy function representing the sum of all the interaction energy of each site with its neighborhood. We denote g to be the spatial and spectral realizations of G, which is in contrast to the spatial-only realizations of the simpler G discussed in [34]. Specifically, g_s are the spectral observations at site s, and g are the collections of these observations for all the sites. Equation (10) is the global distribution for the entire image.

An appropriate localized form of U at a given site s, $U_s(x^P, g)$, is of practical value. The advantage of the Gibbs form is that we can compute the maximum *a posteriori* (MAP) estimate that maximizes the probability $\Pr(x^P | g)$ by iteratively sampling from the local Gibbs distribution pertaining to each site $s \in S_P^{(\sigma)}$. We use

$$\Pr\left(X_s^P = x_s^P | X_r^P = x_r^P, r \neq s\right)$$
$$= \Pr\left(X_s^P = x_s^P | X_r^P = x_r^P, r \in \mathcal{G}_s^{P(\sigma)}\right)$$
$$= \frac{1}{Z_s} e^{-\frac{1}{T} U_s(x^P, g)} \tag{11}$$

where $Z_s = \sum_{x_s^P \in \Gamma} e^{-(1/T)U_s(x^P, g)}$, and $U_s(x^P, g)$ is the energy interaction of site $s \in S_P^{(\sigma)}$ with the neighborhood $\mathcal{G}_s^{P(\sigma)}$. Quite conveniently, the computations of Z_s are not necessary for purposes of obtaining a MAP estimate.

The result of the partition process is a set of partitions (decomposing the set $S_P^{(\sigma)}$)

$$\Xi = \left\{ \Xi_q, q \in \Gamma, \Xi_q \subset S_P^{(\sigma)} \right\} \tag{12}$$

where Ξ_q denotes the collection of all s in $S_P^{(\sigma)}$, for which random variable X_s^P takes the value q. For each q, there will be a set, their union forming a decomposition of $S_P^{(\sigma)}$.

B. Maximization of $L(\beta | X^F, X_*^P)$

We define

$$H = \{h^{(1)}(\lambda), \ldots, h^{(N_{\text{ends}})}(\lambda), \lambda \in \Lambda\} \tag{13}$$

as the set of material spectra, and $H_s \subset H$ as the set of material spectra at pixel site s. At a site s, $h^{(k)} = (h^{(k)}(\lambda_l), \lambda_l \in \Lambda)$ is the spectra for the kth endmember if $h^{(k)} \in H_s$. In the case where there is one partition, H_s equals H. However, in the case where there are multiple partitions, H_s may be strictly contained in H. N_{ends} is the total number of fundamental materials. Recall that, unless we have accurate ground-truth information about the materials in a scene, the task of determining this number is not trivial, as discussed in Section I-B. The goal is to extract the fundamental materials, called endmembers, and associated compositions for materials imaged in a scene that cannot be observed directly, resulting in improved accuracy over existing isolated full-pixel and mixed-pixel methods.

X^F is defined to be a wavelength-dependent random process governed by the spectral mixing relation

$$X_s^F(\lambda) = \sum_{k:h^{(k)} \in H_s} \beta_s^{(k)} \cdot h^{(k)}(\lambda) + \eta_s(\lambda). \tag{14}$$

$\beta_s = \left(\beta_s^{(k)}, k:h^{(k)} \in H_s\right)$ is the vector of proportions at a site s on a label lattice. The function $\eta_s(\lambda)$ is associated with error due to variance in the endmember spectra. $\hat{\beta}_s^{(k)}$ are the estimates of $\beta_s^{(k)}$ obtained through some type of constrained least squares or other means. Because the hyperspectral cube G generates discrete samples of the spectrum at specific wavelength intervals, we can approximate the SMP process using standard linear modeling matrix notation

$$X_s^F = H_s \beta_s + \eta_s, \quad s \in S_I. \tag{15}$$

The basic SMA in (1) is the special case of there being only one partition, i.e., (14), as is, without the modification for the partitions.

The physics of the spectral mixing phenomenon implies that the estimate of β_s should incorporate two constraints: a strict positivity constraining $\beta_s^{(k)} > 0$ for $k:h^{(k)} \in H_s$ and a sum to unity constraint $\sum_{k:h^{(k)} \in H_s} \beta_s^{(k)} = 1$. These are properties exhibited by compositional data.

Therefore, care should really be taken to treat the $\hat{\beta}_s$ as compositional data. Statistical tools for analyzing compositional data have been developed [36]; use of these tools is envisioned in a more sophisticated process than the current one, as will be discussed at the end of this section.

C. Construction of the $\Pr(X^P | G)$

Now suppose we are working with high-quality data and have taken proper precautions so that we can impose the assumption of uncorrupted data such that (3) can be written simply as $G = X^F$. Roughly speaking, without loss of generality, we are thinking of all the noise as being contained in the η term of (14).

We can speak of the full data as being (2), i.e., (X^F, X^P). What is observed is not the full data, but rather only X^F. In principle, one could obtain the marginal probability density $\Pr(X^F|\beta)$ by integrating out X^P (from the joint density $\Pr(X^F, X^P|\beta)$). One could then maximize the likelihood $L(\beta|X^F)$.

Because of the enormous inherent complications of such a complex integration (over all possible realizable partitions), we use the alternative two-step procedure, motivated by the expectation maximization (EM) algorithm [37]

Step 1) Calculate the MAP estimate of X^P, denoted as X^P_*, (analogous to the expectation step)

$$X^P_* = \arg\max \, \Pr(X^P|X^F). \tag{16}$$

Step 2) Calculate the maximum likelihood estimate (MLE) of β, denoted as $\hat{\beta}$, (the maximization step)

$$\hat{\beta} = \arg\max \, L(\beta|X^F, X^P_*) \tag{17}$$

using the MAP estimate of X^P as if it were "observed."

This is analogous to the EM algorithm, except since $\Pr(X^P|X^F)$ does not depend on β in our case, we do not need to iterate as in the usual EM procedure.

D. Extension of the Model

Ultimately, a more sophisticated process is envisioned, namely

$$X = (X^F, X^C, X^L, X^P) \tag{18}$$

where the components of X are a hierarchy of MRFs. X^L would be a stochastic line process identifying a boundary that is defined as an MRF with respect to a graph structure G^L that defines its neighborhood system. Such line elements roughly "form" the boundaries separating the partitioned regions. X^C would be a stochastic compositional process that further constrains the end-member combinations and governs a relaxation of the mixing proportions $\beta_s^{(k)}$ within local neighborhoods. There is a rich statistical subject area concerning compositional analysis that can be used to develop techniques for determining this compositional process [36]. The line process is envisioned to provide the capability to sever neighborhood structures between sites belonging to different compositional phenomenon. Our goal is formalized as the ability to simulate from $\Pr(X = x|G = g)$, via probabilistic relaxations in stages, i.e., $(X^P|G), \ldots$, etc., with the ultimate goal of replacing the process in (2) with the more sophisticated process in (18). The end result is a restoration of the hyperspectral cube from which the local materials and their compositions can be easily extracted.

III. DESCRIPTION OF ALGORITHMS

A. Partition Algorithm

The Gibbs-based algorithm implemented here is based on the framework just discussed. However, before one can implement the algorithm, a few practical matters have to be considered. Specifically in the following A1-A3, we define the energy function of the Gibbs distribution introduced in (11), the associated

neighborhood structures G_s, and the method of computation to maximize the probability in (10).

The algorithm was designed to respond to differences in the shapes of spectral curves and to not be unduly sensitive to illumination differences. Consequently, the energy function below, measures the disparities in spectral angle between sites and their neighbors. This is in contrast to the traditional clustering algorithms, such as KMEANS and ISODATA, that use the Euclidean distance metric as a disparity metric (as well as a significantly different approach).

1) *Energy Function:* The energy function used is

$$U_s\left(x^P, g\right) = \sum_{r \in G_s} \delta\left(x_s^P, x_r^P\right) \cdot f(\theta_{r,s}) \tag{19}$$

$$\text{where } \theta_{r,s} = \cos^{-1}\left(\frac{g_s \cdot g_r}{|g_s| \cdot |g_r|}\right). \tag{20}$$

The delta function $\delta(x_i, x_j) = 1$ for $x_i = x_j$, and $\delta(x_i, x_j) = 0$ for $x_i \neq x_j$. The quantity $\theta_{r,s}$ is the spectral angle between s and r. If $\theta_{r,s} \leq K_{\text{thresh}}$, then $f(\theta_{r,s}) = -1$, otherwise $f(\theta_{r,s}) = 1$, where the threshold K_{thresh} determines the sensitivity of the algorithm to differences in spectral angle. The effect of this energy function is to accumulate the value of $f(\theta_{r,s})$ whenever the label of a site s is equal to its neighbor at r.

The value of the threshold K_{thresh} is determined through observation, specifically, it is determined by observing the differences in spectral angle for samples of the same material versus the differences in spectral angle for samples of different materials. If the imagery to be processed is calibrated to reflectance, then the value attained should be fairly universal, presuming a representative set of materials has been analyzed. If the imagery has not been calibrated, then this value will depend on the range and scaling of the uncalibrated data. A proposed value for calibrated data is given in Section V based on observations of the spectral angles between the numerous spectra in a reference spectral database.

Conceptually, one can think of the energy function as an accumulation of attractive and repulsive interactions between sites and neighbors. Whenever a site and its neighbor are in a common state (i.e., they are labeled the same), they experience an attraction for $\theta_{r,s} \leq K_{\text{thresh}}$ and they experience a repulsion for $\theta_{r,s} > K_{\text{thresh}}$. The attractive interaction contributes negative energy and the repulsive interaction contributes positive energy. If a site and its neighbor are not in a common state, they do not interact.

2) *Neighborhood System:* As indicated from the beginning, we implement a hierarchical multigrid approach. Therefore, the neighborhood systems change according to the grid resolution σ. The algorithm proceeds in a sequence $\sigma = \sigma_{\text{max}}, \ldots, 8, 4, 2, 1$. However, throughout the sampling, we use a predetermined neighborhood structure that satisfies the condition that $s \in G_s$ and $s \notin G_r$ if and only if $r \in G_s$. The structure of the neighborhood system G_s forms a rather simple pattern. A neighborhood of a site will consist of near, intermediate, and far neighbors, and they are all at specific multiples of σ. Our local neighborhood is either the four closest (above, below, left, and right), or eight closest (the surrounding perimeter) sites enclosing the site s. The intermediate and far

neighbors form a similar pattern as the near neighbors but are at a further distance that is a specific multiple of σ. The intermediate and far neighbors are used to aid in obtaining a faster global solution and to enable the algorithm to remember label assignments computed by the algorithm at prior stages.

3) Calculating the MAP Estimate: A simulated annealing technique is used here to obtain the MAP estimate by the Metropolis algorithm as described in [32], [34], and [38] with an eventual "fast freeze" at low temperature. For each iteration, we cycle through all the sites in an image. For a particular site s and iteration k, we compute

$$\gamma_s(k) = \exp\left[-\frac{1}{t_k}[U_s(\tilde{x}^P, g) - U(x^P, g)]\right]. \quad (21)$$

The quantity $\gamma_s(k)$ is the ratio of two Gibbs distributions, specifically, the ratio of the Gibbs distribution for a proposed configuration $\tilde{x}^P$ and the Gibbs distribution of the current configuration x^P, where the Z_s cancels. The proposed configuration $\tilde{x}^P$ is identical to that of x^P, except for a proposal label change at site s. If $\gamma_s(k) > 1$, then a transition is always made (the configuration of x^P at iteration $k + 1$ will be set to the proposed configuration). If $\gamma_s(k) \leq 1$, then the transition is made with a probability of $\gamma_s(k)$. This event occurs whenever $\gamma_s(k) > \xi$, where ξ is a threshold probability that is chosen uniformly in the interval $(0,1)$. Otherwise, no transition occurs (the configuration of x^P at iteration $k + 1$ will remain unchanged). The temperature t_k defines an annealing schedule and we use the one proposed in [32] and [34], namely

$$t_k = \frac{C}{\ln(1 + k)}. \quad (22)$$

C is a parameter corresponding to the maximum depth of a "well" in the energy function (19), which is discussed further in [32]. The value of C affects the rate of convergence. In our case, an upper bound on this value can be computed as the value of (19), for which the spectral angle of the site-neighbor pairs are all different, but the labels are assigned the same value. A large value of C starts the computations at a high temperature, and it is theoretically appealing because a sufficiently high temperature guarantees convergence to the true Gibbs distribution (a global solution). Computationally, a large C is undesirable because it results in very slow convergence and long computation times. Selecting a small value of C tends to cool the system too quickly and causes the algorithm to get stuck in a local minimum, where convergence to the true Gibbs distribution is not achieved. Because we are using a multigrid approach, the computing time can be reduced considerably by using larger values for C at the coarser grids, but smaller values at the finer grids. Practical values of C are discussed in Sections IV and V.

B. Spectral Mixture Process Conditioned by Partitioning (SMPCP)

Basic unconditioned spectral mixture analysis is the special case where the scene is treated as having one partition. In this one partition, there is one set of endmember spectra so that

$H_s = H$, $s \in S_I$. There are obvious disadvantages to this approach. First, the estimates $\hat{\beta}_s^{(k)}$ for (15), provided by a least squares form of (17), constrained or unconstrained, are most stable when there are only a small number of endmembers in the model. For typical scenes, there are often many. Using a small number of endmember spectra for such scenes results in large residual errors (poor estimates) for many sites, whereas using the larger full set of endmember spectra results in unstable estimates. Another disadvantage is that there are certain bland spectra, such as asphalt, water, shade, and metal that are very difficult to unravel in a mixture model, and lumping them together in a model results in very poor results.

With multiple partitions we can assign a different set of endmember spectra to each partition within the framework described in Section II-B. This method will be called "spectral mixture process conditioned by partitioning" (SMPCP), and it is consistent with the framework defined by (12)–(17). Ideally, we would like to determine these different endmember spectra automatically through methods such as [24]–[29]. However, as mentioned in the introduction, this is a very ambitious task. In an effort to control the sources of error, we extract image-derived endmember spectra, and based on known ground-truth of the scene, manually assign these endmembers to partitions.

To accomplish the conditioning, partition masks are constructed using the output class maps generated by the Gibbs-based algorithm. A set of permissible endmember spectra is assigned to each partition to obtain H_s. Estimates of $\hat{\beta}_s^{(k)}$ in (15) are computed using constrained spectral mixing, where the masks are used to determine which set of endmember spectra are permissible in the model. The spectral toolbox available in ENVI [39], which applies the sum to unity constraint, is used to compute the $\hat{\beta}_s^{(k)}$ estimates on the masked data in a sequence of mutually exclusive operations.

IV. DESCRIPTION OF EXPERIMENT

A. Data

Two HYDICE scenes extracted from a data run collected at 10 000 feet altitude near Fort Hood, TX, on May 16, 1996, are used in the experiment. Each scene consists of 300 samples × 320 lines of pixel vector data with a spatial resolution of approximately 2 m. HYDICE Band 49 of Scenes 1 and 2 are shown in Figs. 1 and 2, respectively. Complete HYDICE hyperspectral image cubes have 210 bands in the approximate 400–2500 nm spectral range. However, because of the overwhelming degradations that occur in the atmospheric windows and occasion line degradations occurring in certain bands, only 117 bands were actually used in the analysis. Fig. 3 shows the regions (shaded) that were excluded from the processing, as well as grass and soil spectra extracted from Scene 2. The 117 bands of data actually used in the computations, for each scene, were calibrated to reflectance by the empirical line method based on the known spectra of the calibration panels that are evident in Scene 2.

Ground truth, documenting the identity of some of the ground features in Scene 1, is based on data collected during a site visit to the area during April 2000. As part of our experiment, the data from this site visit (a separate effort) were compiled as a
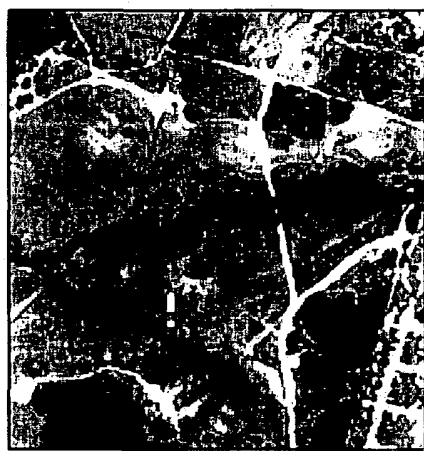
Fig. 1.   Scene 1, HYDICE Band 49.

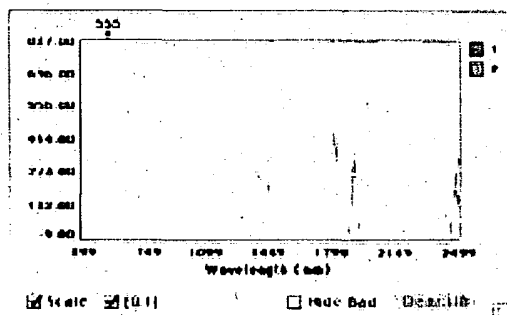Fig. 2.   Scene 2, HYDICE Band 49.

Fig. 3.   Spectral signatures of grass (#1) and soil (#2).

Fig. 4.   Ground truth overlay of Scene 1.

TABLE I
DESCRIPTION OF THE GROUND TRUTH (GT), AS DISPLAYED IN FIG. 4

GT Label	Description of the Area
A	Asphalt parking lot, newly paved.
B	Rooftop of a department store, light-gray asphalt shingles, skylights, yellow gas pipes.
C	Concrete pavement to loading dock, some tire marks.
D	Asphalt parking lot.
E	Church and school with metal rooftop.
F	Private residence, stand-alone.
G	School, rooftop coated with bubbly light-gray material, top edge is gravel.
H	Apartment buildings with asphalt-shingled rooftops.
I	Corrugated steel buildings.
J	Shade of building.
K	Asphalt parking lot surrounding an apartment complex.
L	Church with dark brown asphalt-shingled rooftop.
M	Tennis court, concrete with blue/gray coating, and lots of dirt.
N	Playground area with exposed soil.
O	Encircled area is residential with moderately-priced homes and concrete driveways.
P	Asphalt road network within encircled area.
Q	Light linear pattern is exposed bedrock.
R	Trees - Deciduous holly, Bois D'Arc, Osage Orange, Texas Oak, and Bow Wood
S	Trees - Red Bud, Texas Oak, and Red Bud
T	Scattered junipers throughout the grassy areas.
U	Tall wild grass.
V	Asphalt road intersection with exposed soil along the shoulders.
W	Playground with exposed soil, gravel, and some asphalt.
X	Tall wild grass.
Y	Exposed soil and bedrock.

experiment at other geographic locations, were used to determine an appropriate quantization threshold for the spectral angle disparity metric, as discussed in Section IV-B. Unfortunately, ground-measured spectral data within the study site were not available.

Of the two scenes, Scene 1 has the type of complexity where it is more appropriate to apply the entire Gibbs-based conditioned spectral mixture analysis. Due to the nature of Scene 2 (almost exclusively vegetation), this scene does not stand to benefit much from a spectral mixing process conditioned by multiple partitions. In the context of conditioning the spectral mixing process, one partition is sufficient. Scene 2 is really an example of the kind of scene that is appropriate for either traditional SMA or traditional full-pixel methods. In our experiments, Scene 1 was used to test the entire process. Scene 2 was

graphic overlay shown in Fig. 4, with an associated descriptive table, Table I. Because of the time difference between scene acquisition and the site visit, not all portions of the scene could be documented with certainty. In addition to the ground-feature data, spectral data of a diverse range of materials, measured by field spectrometers and collected independently of this

Fig. 5. ISODATA cluster map of Scene 1.

used primarily to test the partition algorithm in a different scene environment.

B. Characterization and Validation of the Partitioning Algorithm

Computer code was written in C++ to implement the Gibbs-based algorithm in the manner described in Sections II and III. The spectral angle disparity metric was implemented as defined in (19) and (20). In order to determine an appropriate value of the quantization threshold K_{thresh}, the spectral angles between all the pairwise combinations of the spectra in the reference spectral library, just mentioned previously, were computed.

The two scenes were processed in multigrid sequences of runs. Because the algorithm is new, some experimentation with the parameters is necessary to get a flavor of the algorithm's behavior with different settings. The parameters that affect this behavior are σ (grid resolution), C (controls the annealing schedule through the temperature t_k), K_{total} (final number of iterations through the image), and N_{labels} (the number of partition labels). After some initial probing of the parameters, three trials (Trials A, B, and C) were performed on Scene 1 and two trials (Trials D and E) were performed on Scene 2. The results are documented in Section V, which include the computation times for the runs.

The unsupervised ISODATA clustering method is used as a reference method, and serves to provide motivation for using the Gibbs-based approach. Fig. 5 shows the results of this algorithm on Scene 1, for 6 classes.

C. Spectral Mixing Experiment

Two spectral mixing trials were performed using Scene 1, which is considered the more appropriate of the two scenes for the conditioned spectral mixture analysis. In the first trial, conventional spectral mixture analysis was performed in the conventional manner with the ENVI spectral toolbox. Ground truth was used to determine appropriate endmembers. The spectral data of the endmembers were extracted from the scene, based on known locations of the endmembers. The class names are listed in the first column of Table II.

TABLE II
LIST OF ENDMEMBER CLASSES USED IN THE SPECTRAL MIXING TRIALS

No Partitioning	Partition A	Partition B
Bright Roof (B)	Bright Roof (B)	Concrete/ Exposed Soil (Y)
Concrete/ Exposed Soil (Y)	Concrete/ Exposed Soil (Y)	Trees (R)
Asphalt Pavement (A)	Asphalt Pavement (A)	Grass (X)
Asphalt Shingles (H)	Asphalt Shingles (H)	Shade (J)
Trees (R)	Shade (J)	
Grass (X)	Grass (X)	
Shade (J)		

TABLE III
CONTROL PARAMETERS FOR THE PARTITIONING TRIALS. THE CONTROL PARAMETERS FOR THE SCENE 1 TRIALS (TRIALS A, B, AND C) AND THE SCENE 2 TRIALS (TRIALS D AND E) ARE SHOWN BELOW. FOR EACH GRID RESOLUTION σ, WE LIST THE PARAMETER C, THE NUMBER OF ANNEALING ITERATIONS, THE NEIGHBORHOOD STRUCTURE (THE PIXEL DISTANCE OF THE NEIGHBORS FROM A SITE), AND THE RUNNING TIME ON A 500 MHz LAPTOP COMPUTER. (A) LIST OF CONTROL PARAMETERS FOR TRIAL A. (B) LIST OF CONTROL PARAMETERS FOR TRIAL B. (C) LIST OF CONTROL PARAMETERS FOR TRIALS C AND D. (D) LIST OF CONTROL PARAMETERS FOR TRIAL E

σ	C	No. of Iterations	Neigh. System	Run Time
8	16	150,000	8, 32, 64	018 min. 05 sec.
4	8	50,000	4, 16, 32	024 min. 30 sec.
2	4	50,000	2, 16, 32	096 min. 09 sec.
1	4	50,000	1, 16, 32	386 min. 25 sec.

(a)

σ	C	No. of Iterations	Neigh. System	Run Time
8	16	150,000	8, 32, 64	18 min. 05 sec.
4	8	50,000	4, 16, 32	24 min. 30 sec.
2	1	3,000	2, 16, 32	06 min. 20 sec.
1	1	1,000	1, 16, 32	09 min. 07 sec.

(a)

σ	C	No. of Iterations	Neigh. System	Run Time
8	16	150,000	8, 32, 64	18 min. 05 sec.
4	1	2,000	4, 32, 64	01 min. 23 sec.
2	1	700	2, 4, 8	01 min. 52 sec.
1	1	700	1, 4, 8	06 min. 43 sec.

(a)

σ	C	No. of Iterations	Neigh. System	Run Time
16	9	50,000	16, 32, 64	02 min. 45 sec.
8	3	2,000	8, 32, 64	00 min. 35 sec.
4	3	2,000	4, 32, 64	01 min. 24 sec.
2	1	700	2, 4, 8	01 min. 55 sec.
1	1	350	1, 4, 8	04 min. 00 sec.

(a)

In the second mixing trial, a simplified approach of conditioning the spectral mixing process on the partition process was applied. Ground truth was used to aid in the assignment of endmembers to the partitions that were generated in Trial A by the Gibbs-based algorithm. Based on scene knowledge and the partition results, the six partitions resulting from Trial A were collapsed into two partitions (A and B), representing two types of phenomenon: natural/vegetation and cultural/building materials. The endmember spectra extracted in the first mixing trial were reassigned so as to belong to Partition A or Partition B. These assignments are listed in the right two columns of Table II. Note that although the partitions are mutually exclusive, the class assignments often overlap. Also, note that the letters in parentheses of Table II indicate the regions of the scene where the spectra were extracted, as shown in Fig. 4 and Table I.

V. RESULTS

Fig. 5 and Table IV(a) show the results of the ISODATA clustering method applied to Scene 1 for six classes, using the same bands selected for the Gibbs-based sequences discussed below. Although ISODATA provides a fairly good partitioning of some materials in the scene, it has a very bad problem confusing asphalt with significant portions of vegetation in the forested
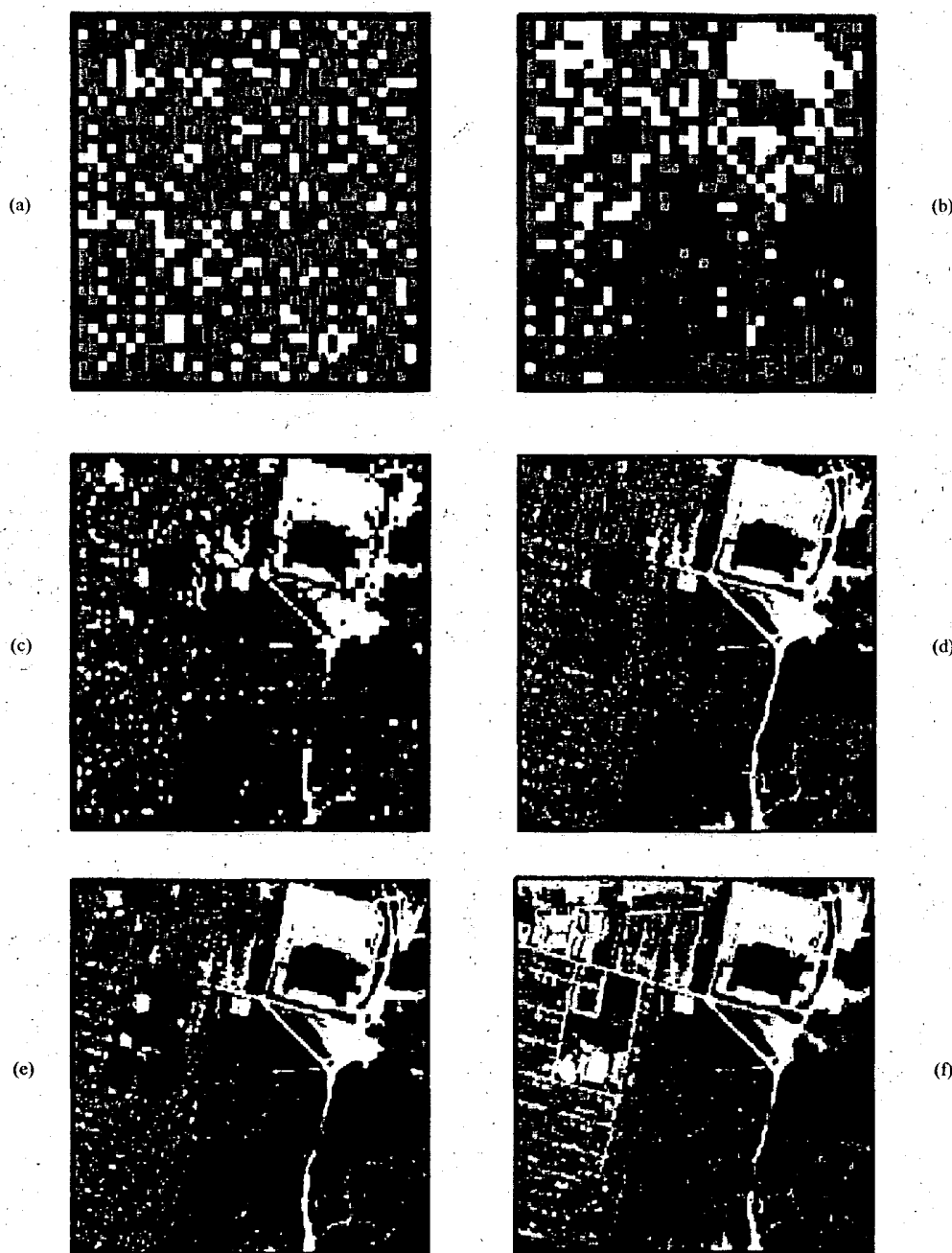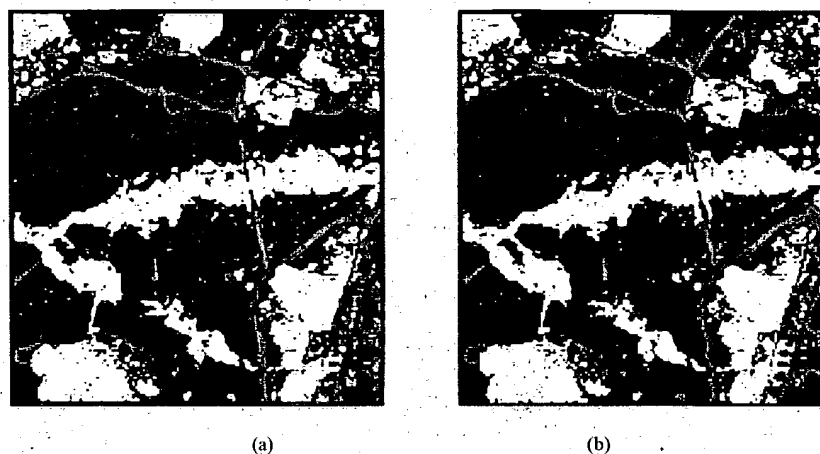
Fig. 6.   Partitioning results for spectral angle disparity, Scene 1. (a) Random initialization $\sigma = 8$, (b) results for trial A $\sigma = 8$, (c) results for trial A $\sigma = 4$, (d) results for trial A $\sigma = 1$, (e) results for trial A $\sigma = 1$, (f) results for trial B $\sigma = 8$, and (f) results for trial C $\sigma = 1$.

area, denoted by "R" in the ground-truth map (Fig. 4), and has difficulty distinguishing between grass and the trees in "R." These problems are seen explicitly in Table IV(a). Out of the 3019 pixels extracted from the "R" region, 580 were placed into ISO-01, 1426 were placed into ISO-02, and 1012 pixels were placed into ISO-04. The cluster associated with ISO-04 can be associated with asphalt by observing Table IV(a), Row A. The cluster label ISO-02 can be associated with Grass by observing Table IV(a), Row U. For purposes of the subsequent conditioned spectral mixture process, confusion between grass

and trees is not problematic. However, confusion between asphalt and (slightly shaded) vegetation is very problematic. The clustering also does not provide a smooth mapping of regions, spatially, and exhibits a significant amount of speckling.

Based on observations of spectral angle distances between the various pairs of materials, we set $K_{thresh} = 11.0$, which remained constant throughout the partition trials. In most of the pairwise signature combinations, there was an obvious gap between similar materials and different materials, except for combinations involving a vegetation signature. Unfortunately, veg-

(a)   (b)

Fig. 7.   Partitioning results for spectral angle disparity, Scene 2. (a) results for trial D $\sigma = 1$, and (b) results for trial E $\sigma = 1$.

etation has a fairly large variance which makes the threshold boundary a bit fuzzy. The exact value was observed to not be terribly critical. However, values smaller than $K_{thresh} = 8.0$ would make the algorithm overly sensitive to feature differences, and values larger than $K_{thresh} = 15.0$ would make the algorithm insensitive to feature differences.

The partitioning algorithm was tested for $N_{labels} = 6$ and $N_{labels} = 8$ partition labels, with no significant difference in partitioning capability. Although $N_{labels} = 6$ is obviously too small to account for all the possible materials in a truly global labeling of the scene, it is also obvious that we are unlikely to determine the true number exactly, in an unsupervised mode. This number is usually sufficient, however, to account for any local variations of materials in a geographic area, and, in the case of moderately complex scenes, to capture the predominant materials globally. We should emphasize that the goal of the partitioning algorithm is not to determine the identity of partitions. The identification of the composition in a scene occurs through the conditioned spectral mixing process. Therefore, determining the exact number of partitions is not critical. However, if the algorithm is applied in a supervised manner (an alternative approach that was suggested in Section I), the number of partitions could be increased or decreased according to scene knowledge.

Table III lists the values of parameters used to generate the partition maps shown in Figs. 6 and 7, as well as the computation time that is required on a 500 MHz G3-based laptop computer. The range $C = (1.0-16.0)$ was investigated, which is below the upper bound discussed in Section III-A.3. The specific value affects the rate of convergence. As an example, for $C = 16$, the destiny of the partitioning results seemed to be controlled by the first 150 000 iterations, and for a value for $C = 2$, the destiny of the partitioning results seems set within the first 2000 iterations.

Fig. 6 shows the results of the Gibbs algorithm applied to Scene 1. As indicated in Table III, each trial is executed as a multigrid sequence of runs, proceeding from a course resolution to finer and finer resolutions. The output of one stage is used to initialize the next stage at a finer resolution. Fig. 6(a) shows the random initialization of Trial A. Figs. 6(b)–(d) are the result of Trial A using the parameter values listed in Table III(a). Summing the corresponding run times, the total computing time

TABLE IV
CONFUSION MATRICES OF RESULTS. LABELS IN THE 15 ROWS ARE THE GT LABELS, AND LABELS IN THE SIX COLUMNS ARE THE PARTITIONED LABELS. AS AN EXAMPLE OF HOW TO READ TABLE IV (A), THE REGION OF GROUND TRUTH EXTRACTED FOR L CONSISTED OF 82 PIXELS. OF THE 82 PIXELS, 44 PIXELS WERE ASSIGNED TO CLUSTER LABEL ISO-03, 37 PIXELS WERE ASSIGNED TO CLUSTER LABEL ISO-04, AND ONE PIXEL WAS ASSIGNED TO THE CLUSTER LABEL ISO-06. (A) RESULTS FOR THE ISODATA ALGORITHM AND (B) RESULTS FOR THE GIBBS-BASED ALGORITHM: TRIAL A

Label	ISO-01	ISO-02	ISO-03	ISO-04	ISO-05	ISO-06
A	0	0	0	232	0	0
B	0	0	0	0	163	0
C	0	0	5	0	0	87
D	0	0	27	0	0	0
E	0	0	0	0	3	25
G	0	0	0	0	4	67
H	0	0	0	67	0	0
I	0	0	0	0	1	60
K	0	0	18	0	0	3
L	0	0	44	37	1	0
M	0	0	0	0	36	0
R	580	1426	1	1012	0	0
S	48	161	0	101	0	0
U	0	506	0	0	0	0
W	0	0	50	0	0	0

(a)

Label	LBL-01	LBL-02	LBL-03	LBL-04	LBL-05	LBL-06
A	0	0	0	232	0	0
B	0	0	0	0	163	0
C	0	0	0	92	0	0
D	0	0	0	27	0	0
E	0	0	0	0	0	28
G	0	0	0	0	71	0
H	0	0	67	0	0	0
I	0	61	0	0	0	0
K	0	0	21	0	0	0
L	0	0	48	34	0	0
M	0	0	23	0	0	13
R	152	2800	16	17	13	21
S	12	292	0	0	1	4
U	506	0	0	0	0	0
W	0	0	50	0	0	0

(b)

of the multigrid sequence for Trial A was quite long (8 h, 45 min). Fig. 6(e) is the final result of Trial B using the parameter values in Table III(b). A comparison of Fig. 6(e) and (d) shows that the result of Trial B is almost identical to the result of Trial A. The total sequence was computed in far less time (58 min), with the difference in execution between the two runs occurring only at the last two stages. Fig. 6(f) is the final result of Trial C, achieved in even less time (28 min) using the values in Table III(c). A comparison between Figs. 6(f) and (e) shows some labeling differences. However, the structural information provided by the partitions that will subsequently be used
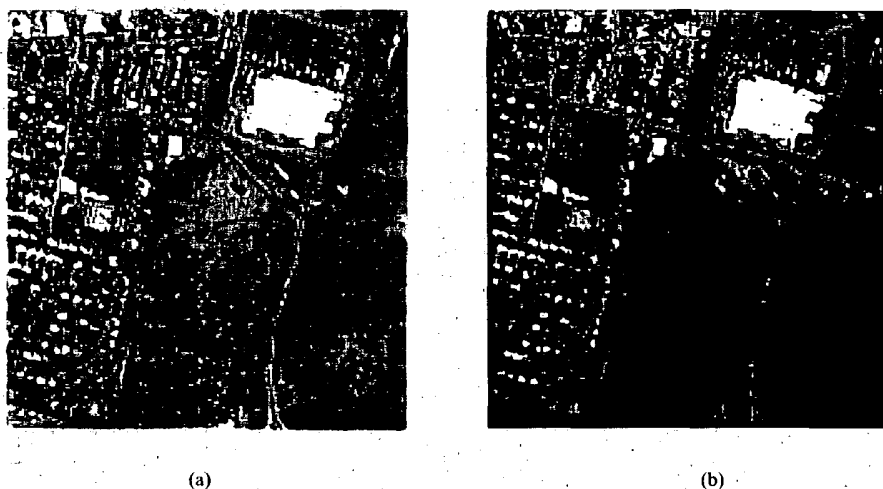
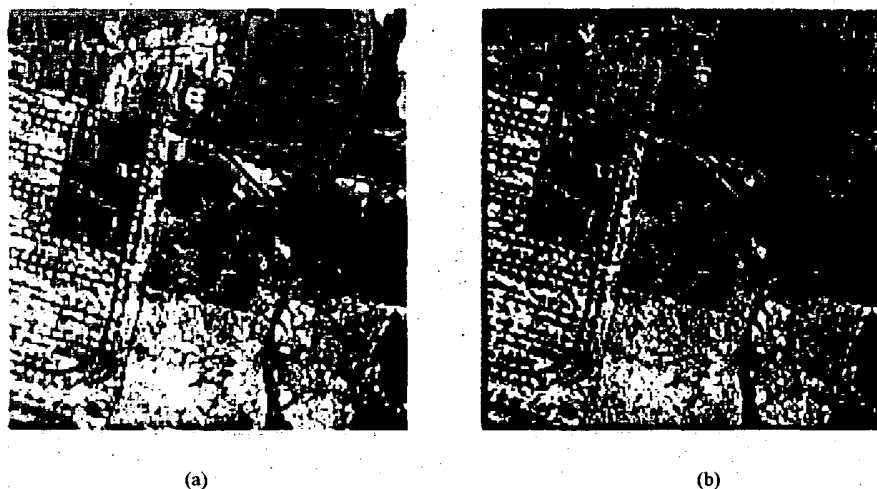Fig. 8.   Abundance maps for bright rooftops. (a) Basic SMA $(X^F|G)$, and (b) SMPCP $(X^F|X^PG)$.

Fig. 9.   Abundance maps for trees. (a) Basic SMA $(X^F|G)$, and (b) SMPCP $(X^F|X^PG)$.

by the conditioned spectral mixing process is not significantly different.

Table IV(b) shows quantitative results for Trial A. A comparison of Tables IV(a) and (b) provides justification that the Gibbs-based partitioning gives advantages with respect to the partition map extracted with ISODATA for the subsequent conditioned spectral mixture process. Observing rows A, C, and D [Table IV(b)], the label LBL-04 can be associated with asphalt or concrete. Observing row U, LBL-01 can be associated with grass. Out of the 3019 pixels extracted from the "R" region, 2800 pixels are assigned a single label (LBL-02), only 17 pixels are labeled as LBL-04, and 152 pixels are labeled as LBL-01. The troublesome confusion between (slightly shaded) vegetation and asphalt that plagues ISODATA is not a problem for the Gibbs algorithm.

Nevertheless, there are still some problems in distinguishing certain urban categories. For example, LBL-03 is assigned to both "H" asphalt-shingle rooftop and "K" asphalt parking lot.

Also, a problem in the global labeling of small and/or skinny features that can be observed in Fig. 6 is not represented in this table. These problems will be addressed in future work.

Fig. 7 shows the results of the Gibbs algorithm for two trials (Trial D and Trial E) applied to Scene 2. Fig. 7(a) is the result of Trial D, running the algorithm with the same parameter values as those used in Trial C. Fig. 7(b) is the result of Trial E, running the algorithm with the values in Table III(d). A comparison of Figs. 7(a) and (b) shows no significant difference in the structural information provided by the partitions and only a few labeling differences. The total run time for Trial E was 10 min and 39 s.

From a theoretical viewpoint, K_{total} should be very large to guarantee a global solution. However, from a practical viewpoint, we would like to reduce the value as much as possible, particularly when the algorithm is operating at a fine resolution $(\sigma = 2, 1)$. Also, from a practical viewpoint, we should keep the starting temperature, controlled by the parameter C, somewhat
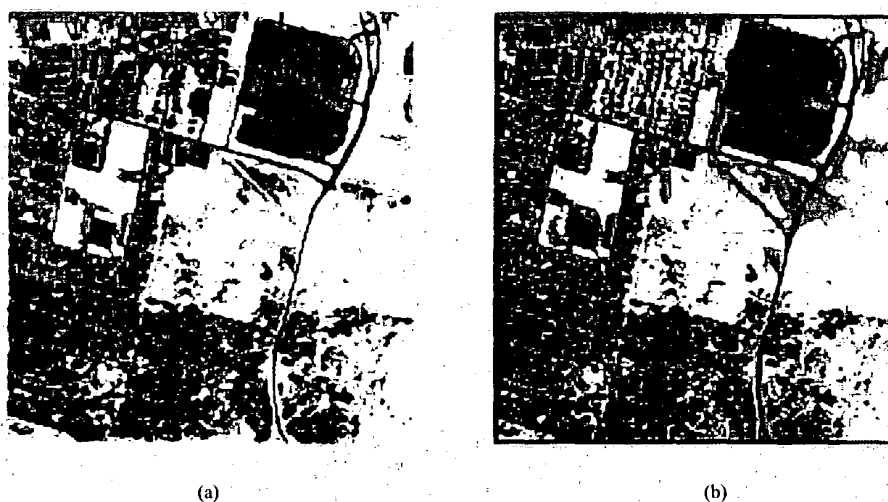
(a)   (b)

Fig. 10. Abundance maps for grass. (a) Basic SMA ($X^F|G$), and (b) SMPCP ($X^F|X^PG$).

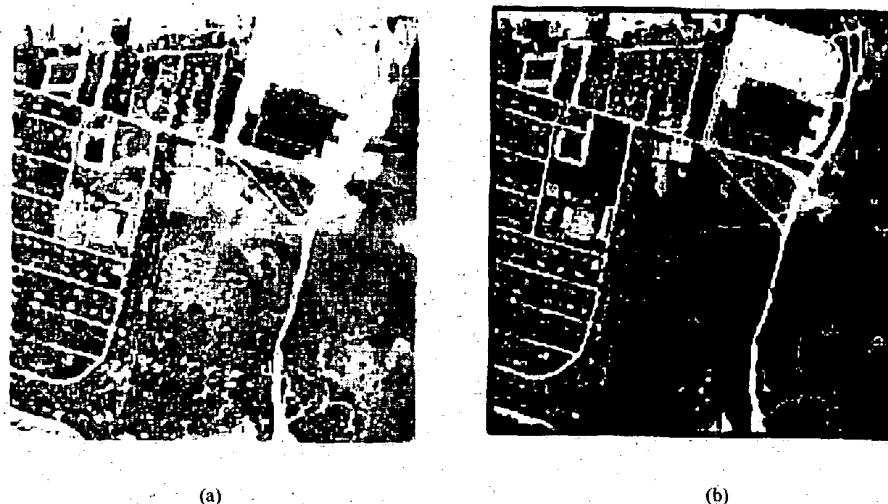(a)   (b)

Fig. 11. Abundance maps for asphalt. (a) Basic SMA ($X^F|G$), and (b) SMPCP ($X^F|X^PG$).

cool at the finer resolutions for two reasons: 1) to reduce the number of iterations and 2) so that the algorithm retains much of the information provided by the coarser resolutions. According to the trials discussed above, Tables III(c) and (d) show a list of parameters that provide practical computing times without adversely affecting the final solution.

Overall, the Gibbs-based algorithm generates partitions that accurately represent structured regions in the image. For all of the bigger regions, pixels with the same label represent the same type of phenomenon on the ground, globally across the image, and regions with different labels correspond to different phenomenon. However, some problems can be seen in a few of the smaller regions. Although the contiguous pixels in smaller partitions represent the same phenomenon, locally, some small partitions, separated by a certain amount of distance and given the same names, actually represent different phenomenon. Conversely, some small partitions, separated by a certain amount of distance and given different names, actually represent similar phenomenon. According to ground truth (Fig. 4), the algorithm

was not able to distinguish the apartment rooftops (H) from the surrounding asphalt parking lot (K). However, note that the rooftop material also contained asphalt (asphalt shingles). We could set the algorithm to have greater sensitivity by lowering the value of the threshold K_{thresh}. However, this would have the undesirable consequence of partitioning the scene into too many regions (particularly, the vegetation), for purposes of the conditioned spectral mixture process.

The problem of global labeling is not unexpected because the Gibbs algorithm, by construction, tends to emphasize local consistency, spatially, by virtue of the MRF formalism. This problem is not as prevalent in algorithms such as KMEANS and ISODATA. However, these other algorithms suffer in another important way because local consistency is ignored. From a perspective of the proposed Bayesian model, the behavior of the Gibbs-based algorithm is preferred, because the ultimate identification of materials in the scene is determined by other model components. Also, recall that we are investigating this approach using a random initialization scheme (in some sense,

a "worst-case scenario," as mentioned in Section I-C). Alternative initialization schemes should substantially improve the global labeling.

The fractional abundance maps on the left side in Figs. 8–11 show the results of performing SMA using one partition with the endmember classes listed in Table II, Column A. The fractional abundance maps on the right side in Figs. 8–11 show the results of performing SMPCP, as described in Section IV-C, using the endmember classes listed in Table II, Columns B and C. Bright areas correspond to an estimated high fraction of the material being represented in the abundance image. Figs. 8–11 are the abundance maps for Bright Rooftops, Trees, Grass, and Asphalt, respectively.

The results show an obvious improvement when using the SMPCP method. The most noticeable improvement is how the unwanted asphalt abundance, prevalent in Fig. 11(a), is reduced by the SMPCP in Fig. 11(b). In the past, the poor estimation of asphalt abundance by basic SMA in regions of shade, vegetation, and water has been a notorious problem. SMPCP offers a practical way to solve this problem. Overall, the SMPCP abundance maps in Figs. 8–11 are sharper, cleaner, and place better emphasis on the material of interest for any specific map.

VI. CONCLUSIONS

The Gibbs-algorithm, using spectral angle as a disparity metric, partitions the imagery into nicely structured regions that are appropriate for input into other subsequent processes in the proposed Bayesian model. The partitions represent homogeneous regions without being overly sensitive to illumination changes on objects in the scene. There were some problems in achieving a global labeling for some of the smaller regions, which should be solvable by using alternative initialization schemes, rather than using random initialization. This problem did not adversely affect the ultimate performance of the combined process.

An improved spectral mixture process was demonstrated using the SMPCP. Partitioning the data provides a mechanism for reducing the dimensionality of the SMP at each site in an image that provides for more stable abundance estimates. Restricting the allowable endmembers in a partition also eliminates false predictions, particularly for endmember spectra that are bland in character, such as asphalt.

The success in applying the Gibb-based and SMPCP algorithms validates the concept of using a Bayesian model and a hierarchy of structure-inducing processes and provides the motivation for developing more sophisticated Bayesian approaches as proposed in Section II-D.

ACKNOWLEDGMENT

The Gibbs-based partitioning algorithm was implemented in C++ code by the authors. However, three other software packages were used to support the effort. HyperCube, developed by R. Pazak, U.S. Army Engineer Research and Development Center (ERDC), Alexandria VA, was used for data manipulation and exploration of the spectral cubes. ENVI, developed by Research Systems, Inc., Boulder, CO, was used for the constrained mixture computations. Multispec, developed at Purdue University, West Lafayette, IN, was used to for the ISODATA classification. The authors would like to thank D. A. Davis, ERDC, who supplied the information needed to construct the ground-truth picture and its corresponding table.

REFERENCES

[1] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, "The Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data," Remote Sens. Environ., vol. 44, pp. 145–163, May–June 1993.

[2] R. Pazak, Hypercube Users Manual. Alexandria, VA: U.S. Army Eng. Res. Devel. Center/TEC, 1999.

[3] S. Bow, Pattern Recognition—Application to Large Data-Set Problem. New York: Marcel Dekker, 1984.

[4] R. Duda and P. Hart, Pattern Classification and Scene Analysis. New York: Wiley, 1973.

[5] C. Therrien, Decision Esimation and Classification—An Introduction to Pattern Recognition and Related Topics. New York: Wiley, 1989.

[6] J. Freeman and D. Skapura, Neural Networks: Algorithms, Applications, and Programming Techniques. Reading, MA: Addison Wesley, 1992.

[7] E. Bosch, "The effects different neural network architectures have on the exploitatin of hyperspectral data," in Proc. Int. Symp. Spectral Sensing Research (ISSSR), Melbourne, Australia, Nov. 1995.

[8] ——, "Classifying multitemportal hyperspectral imagery utilizing neural networks," in Proc. Int. Symp. Spectral Sensing Research (ISSSR), San Diego, CA, Dec. 1997.

[9] R. Rand, "Automated classification of built-up areas using neural networks and subpixel demixing methods on multispectral/hyperspectral data," in Proc. 23rd Annu. Conf. Remote Sensing Soc. (RSS97), Reading, U.K., Sept. 1997.

[10] ——, "Exploitation of hyperspectral data using discriminants and constrained linear subpixel demixing to perform automated material identification," in Proc. Int. Symp. Spectral Sensing Research (ISSSR), Melbourne, Australia, Nov. 1995.

[11] J. Paola and R. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum likelihood classifiers for urban land use classification," IEEE Trans. Geosci. Remote Sensing, vol. 33, pp. 981–996, July 1995.

[12] G. Hepner, "Artificial neural network classification using a minimal training set: comparison to conventional supervised classification," Photogramm. Eng. Remote Sensing, vol. 56, pp. 469–473, Apr. 1990.

[13] J. Adams, M. Smith, and P. Johnson, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 Site," J. Geophys. Res., vol. 91, July 1986.

[14] W. Farrand and J. Harsanyi, "Mineralogic variations in fluvial sediments contaminated by mine tailings as determined from AVIRIS data, Coeur D'Alene River Valley, Idaho," in Proc. 6th Annu. JPL Airborne Geoscience Workshop, Pasadena, CA, 1995.

[15] J. Boardman, "Leveraging the high dimensionality of AVIRIS data for improved subpixel target unmixing and rejection of false positives: Mixture Tuned Matched Filtering," in Proc. 9th Annu. JPL Airborne Geoscience Workshop, Pasadena, CA, 1998.

[16] D. Montgomery and E. Peck, "Introduction to Linear Regression Analysis," in Wiley Series in Probability and Mathematical Statistics, 2 ed. New York: Wiley, 1992.

[17] S. Haykin, Adaptive Filter Theory, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[18] J. Harsanyi and C. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," IEEE Trans. Geosci. Remote Sensing, vol. 32, pp. 779–785, July 1994.

[19] J. Harsanyi, W. Farrand, and C. Chang, "Detection of subpixel signatures in hyperspectral image sequences," in Proc. Amer. Soc. Photogrammetry and Remote Sensing, Reno, NV, 1994, pp. 236–247.

[20] N. Nichols, J. Thomas, W. Kober, S. Johnson, and D. Davis, "Comparative performance of the mixture-tuned matched filter (MTMF) vs. the mixture-tuned matched subspace filter (MTMSF)," in Proc. Int. Symp. Spectral Sensing Research (ISSSR), Las Vegas, NV, Nov. 31, 1999.

[21] A. Green, M. Berman, P. Switzer, and M. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," IEEE Trans. Geosci. Remote Sensing, vol. 26, pp. 65–74, Jan. 1988.

[22] P. Switzer and A. Green, "Min/max autocorrelation factors for multivariate spatial imagery," Dept. Statist., Stanford Univ., Stanford, CA, 1984.

[23] J. Lee, S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote sensing data by a noise-adjusted principal components transform," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 295–304, May 1990.

[24] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Summaries, 4th Annu. JPL Airborne Geoscience Workshop*, vol. 1, Oct. 25–29, 1993, JPL Publ. 93-26.

[25] ——, "Automating linear mixture analysis of imaging spectrometry data," in *Proc. Int. Symp. Spectral Sensing Research (ISSSR)*, San Diego, CA, July 1994.

[26] M. Craig, "Minimum-volume transformations for remotely sensed data," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 542–552, May 1994.

[27] J. Harsanyi, W. Farrand, J. Hejl, and C. Chang, "Automatic identification of spectral endmembers in hyperspectral image sequences," in *Proc. Int. Symp. Spectral Sensing Research (ISSSR)*, San Diego, CA, 1994.

[28] T. Tu, "Unsupervised signature extraction and separation in hyperspectral images: a noise-adjusted fast independent component analysis," *Opt. Eng.*, vol. 39, Apr. 2000.

[29] J. Bayliss, J. Gualtieri, and R. Cromp, "Analyzing hyperspectral data with indepenent component analysis," in *Proc. Applied Image and Pattern Recognition Workshop*. Bellingham, WA, 1997, pp. 133–143.

[30] C. Chang and D. Heinz, "Constrained Subpixel Target Detection for Remotely Sensed Imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 1144–1159, May 2000.

[31] J. Besag, "Toward Bayesian image analysis," *Suppl. J. Appl. Statist.*, vol. 20, no. 5/6, 1993.

[32] G. Winkler, "Image analysis, random fields and dynamic Monte Carlo methods—A mathematical introduction," in *Applications of Mathematics*. Berlin, Germany: Springer-Verlag, 1995.

[33] S. Geman, D. Geman, C. Graffigne, and P. Dong, "Boundary detection by constrained optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 609–628, July 1990.

[34] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Tran. Pattern Anal. Machine Intell.*, vol. PAMI-6, Nov. 1984.

[35] G. Hazel, "Multivariate Gaussian MRF for multispectral scene segmentation and anomaly detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, May 2000.

[36] J. Aitchison, *The Statistical Analysis of Compositional Data*. New York: Chapman & Hall, 1986.

[37] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm (with discussion)," *J. R. Statist. Soc.*, vol. 39, pp. 1–38, 1977.

[38] E. Aarts and Korst, "Simulated Annealing and Bolzmann Machines -A Stochastic Approach to Combinatorial Optimization and Neural Computing," in *Interscience Series in Discrete Mathematics and Optimization*. New York: Wiley, 1989.

[39] *ENVI User's Guide*. Boulder, CO: Research Systems, 1997, vol. 3.2.

Robert S. Rand (M'01) received the B.S. degree in physics from the University of Massachusetts, Lowell, and the M.Eng. and Ph.D. degrees in engineering physics from the University of Virginia, Charlottesville.

He is currently with the U.S. Department of the Army Engineer Research and Development Center (ERDC), U.S. Corps of Engineers, Topographic Engineering Center, Alexandria, VA. His research interests are in enhancing hyperspectral processing techniques through spectral mixture analysis, image segmentation, neural networks, Markov-random field (MRF) models, and independent component analysis.

Daniel M. Keenan received the Ph.D. degree from the University of Chicago, Chicago, IL.

He is currently Associate Professor of statistics with the Department of Statistics, University of Virginia, Charlottesville. His research interests are in applied mathematics, in particular, pattern theory, Bayesian image analysis, and modeling in medicine.

APPENDIX C

A Multigrid Gibbs-based Algorithm to Segment Hyperspectral Imagery using a Combined Spectral Measure of Disparity

Robert S. Rand, *Member, IEEE*, and Daniel M. Keenan

Abstract - *An unsupervised Gibbs-based simulated annealing approach to partitioning hyperspectral imagery into homogeneous regions is investigated, where spatial consistency is imposed on the spectral content of sites in each partition. Maximum A Posteriori (MAP) estimates are computed through the use of a Gibbs distribution defined over a neighborhood system. The algorithm is investigated as a multigrid implementation that uses a non-traditional extended neighborhood system to improve global labeling and reduce computational intensity. An energy function is introduced which models spectral disparities in an image using a combined spectral angle and Euclidean distance measure that is defined with respect to the neighborhood system. The algorithm is focused on terrain mapping applications using hyperspectral imagery containing narrow bands throughout the 400-2500 nanometer spectral region (visible through short-wave infrared).*

Keywords: Bayesian, Markov Random Field, partitioning, clustering, hyperspectral, annealing.

I. INTRODUCTION

We investigate an unsupervised method of segmenting hyperspectral imagery into regions that we shall treat as homogeneous. Our goal is to smoothly label picture elements (pixels) in a scene according to their material type without the use of a reference spectral database or the need for supervised training. Our approach will use the paradigm of Bayesian image analysis.

A mathematical development of the underlying theory, applying the Bayesian framework to Random Processes, is presented in [1]. Bayesian approaches using Markov Random Field (MRF) models have provided algorithms for producing unsupervised partition maps and detecting boundaries on imagery containing one band [2]. Geman *et al* [3] use the Gibbs sampler and Metropolis methods in their algorithms for the restoration of imagery, partititioning, and boundary detection, to compute maximum a posteriori (MAP) estimates. A method to increase the efficiency of global optimization within a general MRF framework for image processing problems has been developed, the Renormalization Group algorithm (RGA), that performs a hierarchical, multiscale, coarse-to-fine analysis of imagery, for the restoration of degraded images [4]. Although multiband imagery has been treated to some extent in the Bayesian community, the

significantly different (spectral) character of hyperspectral imagery has not really been addressed.

In the present effort, we discuss a multigrid Gibbs-based simulated annealing algorithm for partitioning hyperspectral imagery. This effort can be considered an extension, spectrally, of the spatial approach used in [2,3], with additional increased efficiency provided by a multigrid implementation that uses a special case of the general course-to-fine grid structure used in the RGA approach [4]. A non-traditional extended MRF neighborhood system is introduced for improved global labeling.

II. DESCRIPTION OF ALGORITHM

A. Framework

A Bayesian model provides a framework to construct the partitioning of homogeneous regions in hyperspectral imagery. The process X^P, called the "partition process," will identify regions that shall be treated as "homogeneous," in the sense that each region contains some culturally similar phenomenon (not necessarily a single material type). What is actually observed is G, the observed hyperspectral image cube. For G there are both spatial and spectral coordinates.

Our approach for the partitioning is multigrid, and thus we define a collection of lattices that provides a finite sequence of coarser and coarser grids: $S_P^{(\sigma_1)} \to S_P^{(\sigma_2)} \to ... \to S_P^{(\sigma_{N_{seq}})}$, where

$$S_P^{(\sigma_n)} = \left\{ (i\sigma_n + 1, j\sigma_n + 1) : \right.$$
$$\left. 0 \le i \le \tfrac{N_{Rows}-1}{\sigma_n}, 0 \le j \le \tfrac{N_{Cols}-1}{\sigma_n} \right\} \quad (1)$$

N_{Seq} is the number of stages in the sequence, σ_n is a parameter called the "grid resolution" which determines the spatial sampling of the algorithm and the spatial resolution (coarseness) of the labeling process at stage n ($n = 1,...,N_{Seq}$) of the multigrid processing, N_{Rows} is the number of rows in an image, and N_{Cols} is the number of columns. $S_P^{(\sigma_n)}$ is the set of sites that will be used in the partition process at a resolution σ_n. Our grid structure is a special case of the general course-to-fine grid structure used in the RGA approach to image processing [4], where

the coarsest level contains all the large-scale features and each finer level contains correspondingly smaller-scale features. The term "site" refers to a generic element of a lattice, as specified in (1). At full-resolution a site corresponds to a pixel, but at a coarser scale it corresponds to a square block of pixels. The value of σ_n is defined to be the number of pixels per side of each block of pixels. At the finest resolution $\sigma_1 = 1$. We denote the full-resolution image pixel lattice as S_I, i.e., the resolution at which the observations are obtained, where $S_I = S_P^{\sigma_i}$

Now if S is one of $S_P^{(\sigma_n)}$: $S = \{s_1, ..., s_M\}$, where M is the number of sites on the grid, then an associated neighborhood system $G^{(\sigma_n)} = \{G_s, s \in S\}$ is specified (details in Section IIB) which is a set of lattice graph structures of neighbors, such that $s \notin G_s$, and $s \in G_r$ if and only if $r \in G_s$.

X^P is a discrete labeling process which associates a label with each site, and is treated as an MRF. The realizations of X_s^P are $x_s^P \in \Gamma$, where $\Gamma = \{1, 2, ..., N_{Labels}\}$ is the set of partition labels. N_{Labels} is the total number of labels assigned to the partitions. The configuration space for partitioning is then determined by N_{Labels} and σ_n (the label resolution), and denoted by $\Omega_P^{(\sigma_n)} = \{x^P = (x_s^P, s \in S_P^{(\sigma_n)}) \mid x_s^P \in \{1, ..., N_{Labels}\}\}$. The development for X^P is similar to that of [2] with the important addition of a spectral dimension. We treat X^P as a Markov Random Field (MRF) with respect to a graph $\{S_P^{(\sigma_n)}, G^{(\sigma_n)}\}$, where $G^{(\sigma_n)}$ is the entire neighborhood structure for the partition process at resolution σ_n, and G_s is the specific neighborhood at site s for the partition process at resolution σ_n. We exploit the Gibbs equivalence theorem [3] which recognizes the MRF as a random process with a Gibbs distribution. The advantage of adopting this approach is that we can now compute the **maximum a posteriori** (MAP) estimate that maximizes the probability $\Pr(x^P \mid g)$ by iteratively sampling from the local Gibbs distribution pertaining to each site $s \in S_P^{(\sigma_n)}$. We use

$$\Pr(X_s^P = x_s^P \mid X_r^P = x_r^P, r \neq s)$$
$$= \Pr(X_s^P = x_s^P \mid X_r^P = x_r^P, r \in G_s)$$
$$= \frac{1}{Z_s} e^{-\frac{1}{T}U_s(x^P, g)} \qquad (2)$$

where $\{X^P = x^P\}$ means $\{X_{s_1}^P = x_{s_1}^P, ..., X_{s_M}^P = x_{s_M}^P\}$,

$Z_s = \sum_{x^P \in \Gamma} e^{-\frac{1}{T}U_s(x^P, g)}$, and $U_s(x^P; g)$ is the energy

interaction of site $s \in S_P^{(\sigma_n)}$ with the neighborhood G_s. Quite conveniently, the computations of Z_s are not necessary for purposes of obtaining a MAP estimate. We denote g to be the spatial and spectral realizations of G, which is in contrast to the spatial-only realizations of the simpler G discussed in [3]. Specifically, g_s are the spectral observations at site s, and g are the collections of these observations for all the sites.

A simulated annealing technique is used to obtain the MAP estimate by the Metropolis algorithm as described in [1], [3] and [5]. In computing the MAP estimate, we investigate an approach that is consistent with aspects of the Renormalization Group Algorithm (RGA) method developed for the restoration of degraded images [4]. The approach, called "unconstrained simulated annealing," uses a grid-coarsening and cascading technique that is consistent with the RGA approach. Specifically, the fine-to-coarse sequence, described above, with partition lattices defined by (1) is reversed to a coarse-to-fine sequence: $S_P^{(\sigma_{Nseq})} \rightarrow ... \rightarrow S_P^{(\sigma_2)} \rightarrow S_P^{(1)}$. The results of stage "n" are used to initialize the algorithm at the next finer stage "n-1." In the interest of achieving better globalized labeling, we investigate making more than one pass through such a multigrid sequence. A constrained implementation of the algorithm is also currently under development.

B. Neighborhood System

The typical MRF neighborhood is compact (i.e., the closest four or eight neighbors to a site). However, in order to extend the reach of the neighborhood to encourage a more global labeling without a tremendous increase in computation, our algorithm uses a predetermined local, intermediate, and far neighborhood structure that still satisfies the (if and only if) neighborhood condition, mentioned above. Local, intermediate, and far are relative to a given scale (σ_n).

The structure of our neighborhood system G_s forms a simple symmetric pattern. The neighborhood of a site will consist of near, intermediate, and far neighbors, and they are all at specific multiples of σ_n. Our local neighborhood is the eight closest (the surrounding perimeter) sites enclosing the site s. The intermediate and far neighbors form a similar pattern as the near neighbors but are at a further distance that is a specific multiple of σ_n. The intermediate and far neighbors are used to aid in obtaining a faster global solution and to enable the algorithm to remember label assignments computed by the algorithm at prior stages. Of course, different neighborhood structures will produce, when the conditional probability measures (expression (2)) are pieced together, (potentially) different

joint probability measures on the overall configuration space.

C. Energy Function

The energy function, defined below, depends on difference measures, called "disparity measures," which indicate the dissimilarity between a site s and a neighbor r. Our energy function shall combine two disparity measures that measure spectral difference in two very different ways: the spectral angle and the Euclidean distance.

For a pair of sites at the site locations r and s, consider a set $D_{r,s} = \{D_{r,s}^{(1)}, D_{r,s}^{(2)}\}$, where the elements of $D_{r,s}$ are the spectral angle $D_{r,s}^{(1)}$ and the Euclidean distance $D_{r,s}^{(2)}$ between the sites s and r. The individual distance measures $D_{r,s}^{(i)}$, $i=1,2$, are defined as follows:

$$D_{r,s}^{(1)} = \cos^{-1}\left(\frac{g_s \cdot g_r}{|g_s| \cdot |g_r|}\right) \qquad (3)$$

$$D_{r,s}^{(2)} = sqrt\left(\frac{(g_s - g_r)^t \cdot (g_s - g_r)}{N_{bands}}\right) \qquad (4)$$

For the set $D_{r,s} = \{D_{r,s}^{(1)}, D_{r,s}^{(2)}\}$, we define the energy function

$$U_s(x^P, g) = \sum_{r \in G_s} \delta(x_s^P, x_r^P) \cdot \max\{f_i(D_{r,s}^{(i)}), D_{r,s}^{(i)} \in D_{r,s}^{sub}\} \qquad (5)$$

where the delta function $\delta(x_r, x_s) = 1$ for $x_r = x_s$, and $\delta(x_r, x_s) = 0$ for $x_r \neq x_s$. The function $f_i(D_{r,s}^{(i)})$ maps $D_{r,s}^{(i)}$ to the set $\{-1, 1\}$, where $f_i(D_{r,s}^{(i)}) = -1$ indicates a non-significant distance between r and s, and a value $f_i(D_{r,s}^{(i)}) = 1$ indicates a significant distance. If $D_{r,s}^{(i)} \leq K_{Thresh}^{(i)}$ then $f_i(D_{r,s}^{(i)}) = -1$, and otherwise $f_i(D_{r,s}^{(i)}) = 1$, where the threshold $K_{Thresh}^{(i)}$ determines the sensitivity of the algorithm to the $D_{r,s}^{(i)}$ distance measure. The effect of this energy function is to accumulate the value of $f_i(D_{r,s}^{(i)})$ whenever the label of a site s is equal to its neighbor at r.

The value of the threshold $K_{Thresh}^{(i)}$ is determined through observation; specifically, $K_{Thresh}^{(1)}$ is determined by observing the differences in spectral angle for samples of the same material verses the differences in spectral angle for samples of different materials, and $K_{Thresh}^{(2)}$ is determined by observing the differences in Euclidean distance for samples of the same material verses the differences in Euclidean distance for samples of different materials. If the imagery is calibrated to reflectance, then

the value attained should be a fairly universal value, presuming a representative set of materials has been analyzed. If the imagery has not been calibrated, then this value will depend on the range and scaling of the uncalibrated data.

This energy function is designed to respond to disparities as measured by the spectral angle measure and the Euclidean distance. Our reason for using a combined measure is that we seek to increase the sensitivity of the algorithm without introducing undesirable clutter. Rather than using a single measure with a very sensitive threshold $K_{Thresh}^{(i)}$, $i=1,2$, that would respond to one specific type of spectral difference, we use two measures that respond to different kinds of spectral difference with less sensitivity.

III. VALIDATION

A HYDICE scene is used for validation that was extracted from data collected over a geographic area near Fort Hood, Texas at a 10,000 feet altitude, on 16 May 1996. Figure 1 shows the scene, consisting of 300 samples by 600 lines of pixel vector data with a spatial resolution of approximately 1.5 meters. Because of the severe degradations that occur in atmospheric windows of the spectrum and sensor degradations occurring in certain bands, only 117 out of a possible 210 band were used to compute the disparities within the energy functions. These 117 bands of data were calibrated to reflectance by the Empirical Line Method based on the known spectra of the calibration panels that were located in a region outside this particular scene.

Ground truth, describing certain ground features in the scene, was based on data collected during a site visit to the area during April 2000. In addition to the ground-feature data, spectral data of a diverse range of materials, measured by field spectrometers and collected independently of this experiment at other geographic locations, were used to determine an appropriate quantization threshold for the spectral angle and Euclidean distance disparity metrics, as discussed below. Ground-measured spectral data within the study site were not available. Figures 2 and 3 show the results of two trials: an ISODATA clustering run (control method) and a Gibbs unconstrained multigrid run, respectively. Both trials are set to generate six classes. The partitioning algorithm is initialized randomly, and it then proceeds unconstrained in a two-pass multigrid sequence with grid resolutions $\sigma_n = 16,8,4,2 \rightarrow 16,8,4,2,1$. That is, the algorithm proceeds down from the coarsest grid to a grid resolution of $\sigma_2 = 2$, at which point it subsequently steps back again to the coarsest grid and then proceeds to $\sigma_1 = 1$. The distances of the intermediate and far neighbors correspond to fixed multiples of the grid resolution, $4\sigma_n$ and $8\sigma_n$,

respectively. Table I provides commission errors as a very brief summary of the quantitative results. Due to space limitations the original confusion matrices are not shown, and the numerous ground truth (GT) classes are consolidated to three classes.
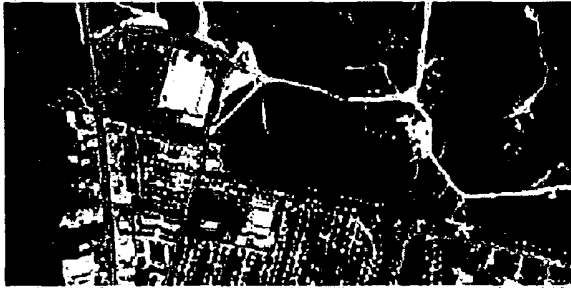

Figure 1. Ft. Hood HYDICE Band 49.

Figure 2. ISODATA Clustering – 6 classes.

Figure 3. Gibbs partitioning – 6 classes.

Table I. Errors of commission.

Commission Errors			
Trial	Trees	Grass	Asphalt
ISODATA	22.48%	1.32%	68.66%
Gibbs	2.60%	2.79%	24.61%

Figure 3 shows a partitioning of the scene that has a significantly smoother appearance, with more well-defined structures, as compared to Figure 2. Quantitatively, Table I shows the Gibbs algorithm provides a significant decrease in the commission errors for Trees and Asphalt.

One negative point is that the Gibbs algorithm (initialized randomly) appears to have a problem labeling small skinny features.

IV. CONCLUSIONS

Overall, the Gibbs-based simulated annealing algorithm generates partitions that accurately represent structured regions in the image with locally smooth labeling. A globally consistent labeling of the partitions was sometimes problematic, particularly, for some of the smaller regions. However, we initialized the algorithm randomly, which is a worst-case scenario. In practice, one might consider alternative initialization strategies. Studies on additional scenes should be performed.

REFERENCES

[1] Winkler G., Image Analysis, Random Fields and Dynamic Monte Carlo Methods – A Mathematical Introduction, Applications of Mathematics, Springer-Verlag, 1995.

[2] Geman S. and Geman D., Graffigne C., and Dong P., "Boundary detection by constrained optimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 7, July 1990.

[3] Geman S. and Geman D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 6, No. 6, November 1984.

[4] Gidas, B., "A Renormalization Group Approach to Image Processing Problems," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2, No. 2, February 1989.

[5] Aarts E. and Korst, Simulated Annealing and Bolzmann Machines - A Stochastic Approach to Combinatorial Optimization and Neural Computing, Interscience Series in Discrete Mathematics and Optimization, John Wiley and Sons, 1989, Reprint1990.

BIOGRAPHY

Robert S. Rand received his Ph.D. in Engineering Physics from the University of Virginia, Charlottesville. He works at the U.S. Department of the Army Engineer Research and Development Center (ERDC), Corps. of Engineers, Topographic Engineering Center, Alexandria, VA.

Daniel M. Keenan received his Ph.D. from the University of Chicago, and is Associate Professor of Statistics in the Department of Statistics at the University of Virginia.

APPENDIX D

Multigrid partitioning of hyperspectral imagery using spectral/spatial measures of disparity

Robert S. Rand, *Member, IEEE*, and Daniel M. Keenan

ABSTRACT

An unsupervised Bayesian approach to partitioning hyperspectral imagery into homogeneous regions is investigated, where spatial consistency is imposed on the spectral content of sites in each partition. An energy function is investigated which models disparities in an image that is defined with respect to a local neighborhood system using one or certain combinations of the spectral angle, Euclidean distance, and/or Kolmogorov-Smirnov (classical and mean-adjusted) measures. Maximum A Posteriori (MAP) estimates are computed using an algorithm that is implemented as a multigrid process to improve global labeling and reduce computational intensity. Both constrained and unconstrained multigrid approaches are considered. A locally extended neighborhood structure is introduced with the intention of encouraging more accurate global labeling. The effort is focused on terrain mapping applications using hyperspectral imagery containing narrow bands throughout the 400-2500 nanometer spectral region, and the trials of our experiment are conducted on a scene from HYDICE 210-band imagery collected over an area that contains a diverse range of terrain features and that is supported with ground truth. Quantitative measures of local consistency (smoothness) and global labeling, along with the class maps, reported in the experiment, demonstrate a clear benefit of this method as compared to the commonly used ISODATA clustering algorithm, where the best results are achieved with a constrained approach using an energy function consisting of a combined spectral angle and Euclidean distance measures.

Index Terms: Markov Random Field, Gibbs distribution, partitioning, clustering, hyperspectral, spectral angle, Kolmogorov Smirnov statistic.

I. INTRODUCTION

The most common unsupervised methods for partitioning multispectral imagery are KMEANS and ISODATA [1], [2], where the measure used in determining the cluster membership is the Euclidean distance. These clustering algorithms are also sometimes used on hyperspectral imagery; however, the Euclidean measure is not particularly responsive to the fine structure, or shapes, in high-resolution spectra, and is often overly sensitive to intensity differences. Furthermore, these methods do not involve any type of spatial localization or consistency operation. A more sophisticated multivariate approach based on the maximum likelihood estimates of first-order and second-order statistics has been developed [3], but relies on dimension reduction methods to operate effectively on the numerous bands inherent with hyperspectral data. This method has been tested using matrix factorization for band selection [4] but alternative dimension reduction methods include Principal Components Analysis [1] and Minimum Noise Fraction [5]. Some related work to [3] investigates a postprocessing method of spatial smoothing using the ECHO classifier that is based on a Markov Random Field (MRF) model [6].

We investigate an unsupervised method of segmenting hyperspectral imagery, where our goal is to smoothly label picture elements (pixels) in a scene according to their material type without the use of a reference spectral database or the need for supervised training. Our approach will use the paradigm of Bayesian image analysis.

An extensive mathematical development of the Bayesian framework for Random Processes and image analysis, including applications, is presented in [7]. Topics relating to Bayesian image analysis are presented in [8], which discusses the Bayesian paradigm, the Gibbs sampler, and methods of estimation. Bayesian approaches using Markov Random Field (MRF) models have provided algorithms for producing unsupervised partition maps and detecting boundaries on imagery containing one band [9]. Geman *et al* [10] use the Gibbs sampler and Metropolis methods in their algorithms for the restoration of imagery, partitioning, and boundary detection, to compute maximum a posteriori (MAP) estimates. A method to increase the efficiency of global optimization within a general MRF framework for image processing problems has been developed by Gidas [11], the Renormalization Group algorithm (RGA), that performs a hierarchical, multiscale, coarse-to-fine analysis of imagery, for the restoration of degraded images.

In prior work [12], we investigated a spectral mixture process that was conditioned by Gibbs-based partitioning. The Gibbs-based method used an MRF model with a simple energy function based on a spectral angle disparity measure, where MAP estimates were computed using an unconstrained multigrid algorithm. The effort showed that significant improvement in spectral unmixing results is achieved on a scene containing many fundamental materials, called "endmembers," if the scene is partitioned into locally smooth regions, where a smaller set of endmembers is attributed to each region.

In the present investigation, we further develop and test the multigrid Gibbs-based segmentation approach. The effort is focused on terrain mapping applications using hyperspectral imagery

containing narrow bands throughout the visible through short-wave infrared spectral region. We

consider both constrained and unconstrained multigrid implementations. We introduce an

"extended" neighborhood structure with the intention of encouraging more accurate global

labeling without sacrificing local smoothness. We also develop and evaluate a number of

spectral/ spatial energy functions. The method can be considered an extension, spectrally, of the

spatial approach used in [9], [10], with additional increased efficiency provided by a multigrid

implementation that uses a special case of the general course-to-fine grid structure used in the

RGA approach [11]. The various algorithm implementations are tested using random

initialization, although, in practice, one can (and probably should) consider some type of

preprocessing method for initialization; we do so presenting a worst-case initial scenario. We

use the ISODATA algorithm as a control method in our experiment, and also use it as a method

of initialization in one of the trials of the experiment. The experiment is conducted on a scene

from HYDICE 210-band imagery collected over an area that contains a diverse range of terrain

features and that is supported with ground truth.

II. MATHEMATICAL FRAMEWORK

A Bayesian model is discussed in this section that provides a framework for the partitioning of homogeneous regions in hyperspectral imagery. The approach is model-based. The process X^P, called the "partition process," will identify regions that shall be treated as "homogeneous," in the sense that each region contains some culturally similar phenomenon (not necessarily a single material type). What is actually observed is G, the observed hyperspectral image cube. For G there are both spatial and spectral coordinates.

The approach will be multigrid, and thus we define a collection of lattices that provides a finite sequence of coarser and coarser grids: $S_P^{(\sigma_1)} \to S_P^{(\sigma_2)} \to \ldots \to S_P^{(\sigma_{N_{Seq}})}$, where

$$S_P^{(\sigma_n)} = \left\{ (i\sigma_n + 1, j\sigma_n + 1): \ 0 \le i \le \tfrac{N_{Rows}-1}{\sigma_n}, \ 0 \le j \le \tfrac{N_{Cols}-1}{\sigma_n} \right\} \tag{1}$$

N_{Seq} is the number of stages in the sequence, σ_n is a parameter called the "grid resolution" which determines the spatial sampling of the algorithm and the spatial resolution (coarseness) of the labeling process at stage n ($n = 1, \ldots, N_{Seq}$) of the multigrid processing, N_{Rows} is the number of rows in an image, and N_{Cols} is the number of columns. $S_P^{(\sigma_n)}$ is the set of sites that will be used in the partition process at a resolution σ_n. Our grid structure is a special case of the general course-to-fine grid structure used in the RGA approach to image processing [11], where the coarsest level contains all the large-scale features and each finer level contains correspondingly smaller-scale features. The term "site" refers to a generic element of a lattice, as specified in (1). At full-resolution a site corresponds to a pixel, but at a coarser scale it corresponds to a square

block of pixels. The value of σ_n is defined to be the number of pixels per side of each block of pixels. At the finest resolution $\sigma_1 = 1$. We denote the full-resolution image pixel lattice as S_f, i.e., the resolution at which the observations are obtained, where $S_f = S_P^{\sigma_1}$.

Let $\Lambda = \{\lambda_l; \lambda_l \in (400\,nm, 2500\,nm), l = 1, N_{Bands}\}$ denote the full-resolution wavelength lattice, where N_{Bands} is the number of spectral bands in the hyperspectral image cube. Hence, for example, the full-resolution coordinates for the observed hyperspectral G are $S_f \times \Lambda : s \in S_f, \lambda \in \Lambda$; therefore, $G(s, \lambda)$ is the observed intensity at site s and wavelength λ.

The above sets, S_f and $S_P^{(\sigma_n)}$, have associated graph structures which describe, for a given element, its neighboring elements. These specified structures allow us to sequentially formulate well-defined probability models for X^P as a Markov Random Field (MRF).

Now if S is one of $S_P^{(\sigma_n)}$: $S = \{s_1, \ldots, s_M\}$, where M is the number of sites on the grid, then an associated neighborhood system $G^{(\sigma_n)} = \{G_s, s \in S\}$ is specified (details in Section IIIA) which is a set of lattice graph structures of neighbors, such that $s \notin G_s$, and $s \in G_r$ if and only if $r \in G_s$. Within this setting, our goal is reduced to finding the value of X^P which maximizes $\Pr(X^P \mid G)$, for the observed G.

X^P is a discrete labeling process which associates a label with each site, and is treated as an MRF. The realizations of X_s^P are $x_s^P \in \Gamma$, where $\Gamma = \{1, 2, \ldots, N_{Labels}\}$ is the set of partition

labels. N_{Labels} is the total number of labels assigned to the partitions. The configuration space

for partitioning is then determined by N_{Labels} and σ_n (the label resolution), and denoted by

$$\Omega_P^{(\sigma_n)} = \left\{ x^P = (x_s^P, \ s \in S_P^{(\sigma_n)}) \mid x_s^P \in \left\{ 1, \ldots, N_{Labels} \right\} \right\}.$$

The development for X^P is similar to that of [9] with the important addition of a spectral

dimension. We treat X^P as a Markov Random Field (MRF) with respect to a graph

$\left\{ S_P^{(\sigma_n)}, G^{(\sigma_n)} \right\}$, where $G^{(\sigma_n)}$ is the entire neighborhood structure for the partition process at

resolution σ_n, and G_s is the specific neighborhood at site s for the partition process at resolution

σ_n. We exploit the Gibbs equivalence theorem [10] that recognizes the MRF as a random

process with a Gibbs distribution. The advantage of adopting this approach is that we can now

compute the **maximum a posteriori** (MAP) estimate that maximizes the probability $\Pr(x^P \mid g)$

by iteratively sampling from the local Gibbs distribution pertaining to each site $s \in S_P^{(\sigma_n)}$. We

use

$$\Pr(X_s^P = x_s^P \mid X_r^P = x_r^P, r \neq s) = \Pr(X_s^P = x_s^P \mid X_r^P = x_r^P, r \in G_s)$$

$$= \frac{1}{Z_s} e^{-\frac{1}{T} U_s(x^P, g)} \tag{2}$$

where $\left\{ X^P = x^P \right\}$ means $\left\{ X_{s_1}^P = x_{s_1}^P, \ldots, X_{s_M}^P = x_{s_M}^P \right\}$, $Z_s = \sum_{x^P \in \Gamma} e^{-\frac{1}{T} U_s(x^P, g)}$, and $U_s(x^P, g)$ is the

energy interaction of site $s \in S_P^{(\sigma_n)}$ with the neighborhood G_s. Quite conveniently, the

computations of Z_s are not necessary for purposes of obtaining a MAP estimate. We denote g to

be the spatial and spectral realizations of G, which is in contrast to the spatial-only realizations of

the simpler G discussed in [10]. Specifically, g_s are the spectral observations at site s, and g are

the collections of these observations for all the sites.

The result of this process is a set of partitions (decomposing the set $S_P^{(\sigma_n)}$):

$$\Xi = \left\{ \Xi_q, \ q \in \Gamma, \ \Xi_q \subset S_P^{(\sigma_n)} \right\},$$ where Ξ_q denotes the collection of all s in $S_P^{(\sigma_n)}$ for which random

variable X_s^P takes the value q. For each q, there will be a set, their union forming a

decomposition of $S_P^{(\sigma_n)}$.

III. DESCRIPTION OF ALGORITHM

Our proposed algorithm is based on the framework just discussed. We implement a hierarchical

multigrid approach, where the algorithm typically proceeds in a sequence $\sigma_n = \sigma_{N_{Seq}}, \ldots, 8, 4, 2, 1$.

We investigate both unconstrained and constrained implementations.

We shall consider energy functions that respond to spectral disparities in an image using spectral

angle and/or Euclidean distance, as well as spatial disparities based on the Kolmogorov Smirnov

statistic as in Geman *et al* [9]. Recall that traditional clustering algorithms, such as KMEANS

and ISODATA, only use the Euclidean distance measure to determine cluster membership.

We now discuss the specifics of the implementation, below, in Sections IIIA-E. We present the

neighborhood system, define a set of disparity measures and a number of specific energy

functions of the local Gibbs distribution defined by (2), as well as a method for computing the

corresponding MAP estimate. We then discuss two methods for proceeding through the stages of the multigrid algorithm. Later, in Section IVC, we define performance measures to characterize the smoothness of the partition labeling.

A. NEIGHBORHOOD SYSTEM

The typical MRF neighborhood consists of the closest four or eight neighbors to a site; below, we will refer to this as the "compact" neighborhood. However, in order to extend the reach of the neighborhood to encourage a more global labeling without a tremendous increase in computation, our algorithm uses an "extended" neighborhood that consists of a predetermined near, intermediate, and far neighborhood structure that still satisfies the (if and only if) neighborhood condition, mentioned above. Near, intermediate, and far are relative to a given scale (σ_n). So, the extended neighborhood is a local neighborhood, but "less so" than the compact one.

The structure of the extended neighborhood system G_s forms a rather simple pattern. The neighborhood of a site will consist of near, intermediate, and far neighbors, and they are all at specific multiples of σ_n. Our local neighborhood is the eight closest (the surrounding perimeter) sites enclosing the site s. The intermediate and far neighbors form a similar pattern as the near neighbors but are at a further distance that is a specific multiple of σ_n. Figure 1 shows such a pattern. The intermediate and far neighbors are used to aid in obtaining a faster global solution and to enable the algorithm to remember label assignments computed by the algorithm at prior stages. Of course, different neighborhood structures will produce, when the conditional probability measures (expression (2)) are pieced together, (potentially) different joint probability

measures on the overall configuration space. More precisely, the resulting probability measure using the "near, intermediate, and far" neighborhood structure is (potentially) a different probability measure from that using only near neighbors. Gibbs sampling and/or simulated annealing are potentially being applied to two different posterior measures. In Section IV, experiments are performed, using HYDICE data, comparing the different neighborhoods. The "extended" neighborhood appears to produce more accurate global labeling, and does so without sacrificing local smoothness.

B. DISPARITY MEASURES

The energy functions that will be defined momentarily depend on difference measures, called "disparity measures," which indicate the dissimilarity between a site s and a neighbor r. We shall investigate disparity measures that measure difference in a number of ways -- specifically, the spectral angle measure, the Euclidean distance, the classical Kolmogorov Smirnov (KS) statistic, and a mean-adjusted KS measure.

For a pair of sites at the site locations r and s, consider a set $D_{r,s} = \{D_{r,s}^{(1)}, D_{r,s}^{(2)}, D_{r,s}^{(3)}, D_{r,s}^{(4)}\}$, where the elements of $D_{r,s}$ are the spectral angle $D_{r,s}^{(1)}$, the Euclidean distance $D_{r,s}^{(2)}$, the classical KS statistic $D_{r,s}^{(3)}$, and a mean-adjusted KS distance $D_{r,s}^{(4)}$, between the sites s and r. The individual distance measures $D_{r,s}^{(i)}$ are defined as follows:

$$D_{r,s}^{(1)} = \cos^{-1}\left(\frac{g_s \cdot g_r}{|g_s| \cdot |g_r|}\right) \qquad (3)$$

$$D_{r,s}^{(2)} = sqrt\left(\frac{(g_s - g_r)^t \cdot (g_s - g_r)}{N_{bands}}\right) \qquad (4)$$

$$D_{r,s}^{(3)} = \sup_{g_j}[\,|\,F_r(g_{r,j}) - F_s(g_{s,j})\,|\,] \qquad (5)$$

$$D_{r,s}^{(4)} = \sup_{g_j}[\,|\,H_r(g_{r,j}) - H_s(g_{s,j})\,|\,] \qquad (6)$$

Equations (5) and (6) need some further discussion. Generally speaking, the KS statistic is used to measure the significance of the difference between two univariate empirical distributions. In our case, the variable $g_{r,j}$ is the response of a specific band j centered at site r in the hyperspectral cube (Note that although we presently just select a single band of imagery, we could easily modify the approach such that we use the mean, the maximum, or some other consolidated measure of a multiple number of bands). $F_r(g_{r,j})$ denotes the empirical distribution function modeling a window of pixels centered at the site r, and $H_r(g_{r,j})$ denotes an adjusted empirical distribution function modeling a window of pixels centered at the site r, where the mean of the window of pixels is subtracted from $F_r(g_{r,j})$. We denote N_{Wind} as the number of pixels in the spatial window. Therefore, $D_{r,s}^{(3)}$ in (5) is the least upper bound of all pointwise differences $|F_r(g_{r,j}) - F_s(g_{s,j})|$. Likewise, $D_{r,s}^{(4)}$ in (6) is the least upper bound of all pointwise differences $|H_r(g_{r,j}) - H_s(g_{s,j})|$.

C. ENERGY FUNCTIONS

For each individual disparity measure in (3) – (6), an energy function is defined as

$$U_s^{(i)}(x^P, g) = \sum_{r \in G_s} \delta(x_s^P, x_r^P) \cdot f_i(D_{r,s}^{(i)}) \quad i=1,4 \tag{7}$$

where the delta function $\delta(x_r, x_s)=1$ for $x_r = x_s$, and $\delta(x_r, x_s)=0$ for $x_r \neq x_s$.

The function $f_i(D_{r,s}^{(i)})$ maps $D_{r,s}^{(i)}$ to the set $\{-1, 1\}$, where $f_i(D_{r,s}^{(i)}) = -1$ indicates a non-significant distance between r and s, and a value $f_i(D_{r,s}^{(i)}) = 1$ indicates a significant distance. If $D_{r,s}^{(i)} \leq K_{Thresh}^{(i)}$, then $f_i(D_{r,s}^{(i)}) = -1$, and otherwise $f_i(D_{r,s}^{(i)}) = 1$, where the threshold $K_{Thresh}^{(i)}$ determines the sensitivity of the algorithm to the $D_{r,s}^{(i)}$ distance measure. The effect of this energy function is to accumulate the value of $f_i(D_{r,s}^{(i)})$ whenever the label of a site s is equal to its neighbor at r.

For $i=1,2$ the value of the threshold $K_{Thresh}^{(i)}$ is determined through observation; specifically, $K_{Thresh}^{(1)}$ is determined by observing the differences in spectral angle for samples of the same material verses the differences in spectral angle for samples of different materials, and $K_{Thresh}^{(2)}$ is determined by observing the differences in Euclidean distance for samples of the same material verses the differences in Euclidean distance for samples of different materials. If the imagery is calibrated to reflectance, then the value attained should be a fairly universal value, presuming a representative set of materials has been analyzed. If the imagery has not been calibrated, then this value will depend on the range and scaling of the uncalibrated data.

For $i=3,4$ a probability of false rejection α can conceivably be set (say $\alpha=.01$) and the threshold could then be determined by analytical values that characterize the KS distribution (tabulated values are available for small windows of $N_{Wind} \leq 45$ pixels, and formulas are designated for larger windows). Applying the criterion in this fashion corresponds to performing the usual hypothesis testing procedure. As will be discussed further in Section IV, this test of hypothesis is too conservative to be useful in practice when applied to the non-adjusted distributions for $i=3$, and experiment observation is again needed to set the threshold.

In addition to the energy functions (7) that respond to a single disparity measure, we can construct an energy function that responds to any pair of disparity measures, or any subset $D_{r,s}^{sub} \subset D_{r,s}$ of the set of disparity measures $D_{r,s}$. For example, we could define $D_{r,s}^{sub1} = \{D_{r,s}^{(1)}, D_{r,s}^{(2)}\}$ as the subset that specifies the spectral angle and Euclidean distance between sites r and s. For such subsets, (7) can be generalized as follows:

$$U_s(x^P, g) = \sum_{r \in G_s} \delta(x_s^P, x_r^P) \cdot \max\{f_i(D_{r,s}^{(i)}), D_{r,s}^{(i)} \in D_{r,s}^{sub}\} \tag{8}$$

Appropriate threshold values for calibrated data are given in Section IV, based on observations for $i=1,2,3$ and based on a statistical test at $\alpha=.01$ for $i=4$.

Conceptually, this energy function is an accumulation of attractive and repulsive interactions between sites and neighbors. Whenever a site and one of its neighbors have identical labels they are in a common state. In this case, they interact, experiencing an attractive interaction for $D_{r,s}^{(i)} \leq K_{Thresh}^{(i)}$ and a repulsive interaction for $D_{r,s}^{(i)} > K_{Thresh}^{(i)}$. The attractive interaction contributes

negative energy to (7) or (8), and the repulsive interaction contributes positive energy. Whenever a site and its neighbor are not in a common state, they do not interact.

D. Calculating the MAP estimate

A simulated annealing technique is used to obtain the MAP estimate by the Metropolis algorithm as described in [7], [10] and [13]. For a particular site s and iteration k, we compute the ratio of two local Gibbs distributions:

$$\gamma_s(k) = e^{-\frac{1}{t_k}[U_s(\tilde{x}^P, g) - U(x^P, g)]} \tag{9}$$

Specifically, $\gamma_s(k)$ is the ratio of the Gibbs distribution for a proposed configuration $\tilde{x}^P$ and the Gibbs distribution of the current configuration x^P, where the Z_s in (2) for each distribution cancels. The proposed configuration $\tilde{x}^P$ is identical to that of x^P, except for a proposal label change at site s. The value of $\gamma_s(k)$ is used to determine whether (or not) a transition occurs, in which case the configuration of x^P at iteration $k+1$ shall be set to the proposed one. A transition is always made whenever $\gamma_s(k) > 1$. If $\gamma_s(k) \leq 1$ then the transition is made with a probability of $\gamma_s(k)$. This latter event occurs whenever $\gamma_s(k) > \xi$, where ξ is a threshold probability that is chosen uniformly in the interval $(0,1)$. Otherwise, no transition occurs (the configuration of x^P at iteration $k+1$ will remain unchanged). An annealing schedule is defined through a temperature t_k. We follow an annealing schedule proposed in [7] and [10], namely,

$$t_k = \frac{C}{\ln(1+k)} \tag{10}$$

The parameter C corresponds to the maximum depth of a "well" in the energy function (7) or (8), and is discussed in [7]. A large value of C starts the computations at a high temperature. A large

C is theoretically appealing because a sufficiently high temperature guarantees convergence to the mode of the Gibbs distribution (a global solution) [10]; however, computationally, a large value is undesirable because it results in very slow convergence and long computation times. Selecting a small value of C tends to cool the system too quickly and causes the algorithm to get stuck in a local minimum, where convergence to the mode of the distribution is not achieved. Because we are using a multigrid approach, the computing time can be reduced considerably by using larger values for C at the coarser grids, but smaller values at the finer grids. Practical values of C are discussed during the experiments in Section IV. Also as a practical matter, once the temperature t_k lowers considerably and a processing stage nears its end, we fast freeze the process by dropping the temperature to zero.

E. MULTIGRID APPROACHES

The multigrid algorithm provides a multi-scale sequence of partitioning with the corresponding site lattices (label images) having different resolution cells. The coarse levels contain the macroscopic features in a scene, whereas the finest level contains all the microscopic features. This is intuitively appealing in the sense that human vision is often thought to work this way. For example, if a person is looking at a painting from a distance he only sees the larger macroscopic features. As he proceeds closer to the painting the smaller microscopic details became apparent. In order to truly understand the picture, the person may need to view multiple scales, and perhaps even step back and forth a few times to "partition" the scene into meaningful patterns.

We investigate two approaches to proceeding through the stages of the multigrid algorithm that are consistent with aspects of the Renormalization Group Algorithm (RGA) method developed by Gidas in [11] for the restoration of degraded images. The first approach denoted as "unconstrained simulated annealing," uses a grid-coarsening and cascading technique that is consistent with one of the techniques used in the RGA method. Specifically, the fine-to-coarse sequence described in Section II with partition lattices defined by (1) is reversed to a coarse-to-fine sequence: $S_P^{(\sigma_{N_{Seq}})} \to \ldots \to S_P^{(\sigma_2)} \to S_P^{(1)}$. The results of stage "$n$" are used to initialize the algorithm at the next finer stage "n-1." However, unlike the RGA method, no constraints are applied.

The second approach denoted as "constrained simulated annealing" uses the same coarse-to-fine grid sequence; however, the following constraint is applied:

$$\mathrm{Pr}^{(n)}\left(x^{P(\sigma_n)} \mid x^{P(\sigma_{n-1})}\right) = \prod_{s \in S_P^{(\sigma_n)}} \delta_{x_s^{P(\sigma_n)} x_s^{P(\sigma_{n-1})}} \tag{11}$$

This constraint permanently fixes the labels assigned on completion at stage n as the algorithm proceeds to stage n-1.

IV DESCRIPTION OF EXPERIMENT

A. DATA

A HYDICE scene is used in the experiment that was extracted from data collected over a geographic area near Fort Hood, Texas at a 10,000 feet altitude, on 16 May 1996. The scene consists of 300 samples by 600 lines of pixel vector data with a spatial resolution of approximately 1.5 meters. Because of the severe degradations that occur in atmospheric windows of the spectrum and sensor degradations occurring in certain bands, a maximum of 117 bands were used to compute the disparities within the energy functions. These 117 bands of data were calibrated to reflectance by the Empirical Line Method based on the known spectra of the calibration panels that were located in a region outside this particular scene.

Ground truth, describing certain ground features in the scene, is based on data collected during a site visit to the area during April 2000. The data for this experiment is compiled as a graphic overlay onto HYDICE Band 49 and is shown in Figure 2, along with associated descriptive information listed in Table I. Because of the time difference between scene acquisition and the site visit, not all portions of the scene could be documented with certainty. In addition to the ground-feature data, spectral data of a diverse range of materials, measured by field spectrometers and collected independently of this experiment at other geographic locations, were used to determine an appropriate quantization threshold for the spectral angle and Euclidean distance disparity metrics, as discussed below. Ground-measured spectral data within the study site were not available.

B. TRIAL PARAMETERS

The experiment consists of ten trials: an ISODATA run, five unconstrained runs that are initialized randomly, one unconstrained run that is initialized by prior ISODATA results, and three constrained runs that are initialized randomly. The ISODATA trial uses the unsupervised ISODATA clustering method with six classes to be used as a reference against which to test the other trials. Table II indicates which of the disparity measures (3) – (6) are used in the energy function, where (7) is used as the energy function for Trials 1, 3, and 4, and (8) is used as the energy function for the remainder of the trials. Trials 2 – 5 are implemented with the parameters shown in Table IIIa. These trials proceed in an unconstrained multigrid sequence with grid resolutions $\sigma_n = 16,8,4,2,1$. The algorithm is initialized randomly and it then proceeds from a grid resolution of $\sigma_5 = 16$ to $\sigma_1 = 1$. As shown in the table, a range $C = (1.0\text{-}9.0)$ was used which is below the upper bound discussed in Section IIID. The value of C was set to its highest value at the initial stage of a run (corresponding to the highest temperature). In addition to the practical computational advantage, a lower C value allows the algorithm to "remember" the results at previous stages. The distances of the intermediate and far neighbors shown in Table IIIa. corresponds to fixed multiples of the grid resolution, $4\sigma_n$ and $8\sigma_n$, respectively.

As indicated by Table II, Trial 6 is initialized by the results of the ISODATA algorithm using the same disparity measures and energy function used in Trial 2.

Trials 7-9 are constrained multigrid runs. These runs are initialized randomly using the same disparity measures and energy function used in Trial 2. As indicated by Table IIIb, the parameters of Trial 7 are almost identical to the parameters in Trial 2. The only differences are

that C is higher for some of the intermediate grid sizes along with an appropriately higher number of iterations. The constraint should allow us to run the algorithm at a somewhat higher temperature (i.e. higher initial C values) without losing the information at coarser grid sizes -- we wanted to test this.

Trial 8 is a constrained run that tests the performance of the algorithm using the typical compact neighborhood system, rather than the extended neighborhood system we propose and have tested in the prior runs. As shown in Table IIIc, Trial 8 is identical to Trial 7 except for the compact neighborhood system. Again, by compact, here we mean the eight nearest neighbors.

Trial 9 is a constrained run that also tests the performance of the algorithm using the typical compact neighborhood system. Because of the anticipated loss in global information using this neighborhood system, we attempt to compensate for the compact neighborhood system by starting the algorithm at a coarser grid size than that used in Trial 8 (or any of the prior trials). Instead of the five stages used in the other trials, this trial is run with six stages using the parameters shown in Table IIId.

C. MEASURES OF PERFORMANCE

Quantitative results are reported for the trials, making specific use of the ground truth (GT) information to generate confusion matrices and to compute a number of quantitative measures of performance. The quantitative measures include global indicators of labeling accuracy such as commission errors (percentage of class labels assigned to a partition that correspond to

categories other than the one attributed to that partition) and omission errors (percentage of GT labels that are assigned to partitions not associated with a specific GT category), and indicators of labeling smoothness such as clutter and speckle to indicate labeling smoothness.

C1. Global Measures

Whereas, errors of omission and commission are widely used to report the performance of supervised classification algorithms and are rather straightforward to compute for these algorithms, the computations of such errors for unsupervised classification are not quite as straightforward. In supervised classification, the identity of the classification labels with respect to ground truth is typically determined in advance and so there is an exact correspondence between the identity of the labels and the ground truth that is built into the confusion matrices. However, in unsupervised classification, the clusters develop independently of the ground truth and the correspondence between the identity of the classification labels and GT labels must be made afterward, and is often not quite so straightforward. In this experiment, a correspondence of the partitioning labels with three GT categories (Trees, Grass, and Asphalt) was attempted. A successful correspondence was usually (but not always) made for assigning three out of the six partitioning labels to these three categories.

C2. Local Measures To Characterize Classification Smoothness

Smoothness in classification labeling can be measured from the viewpoint of clutter and speckle. Suppose we have a region in a scene that is known to be homogenous. Therefore, all the pixels in this region should be assigned to one label. Clutter can be defined as the number of pixels

assigned to labels that are different from the predominant label in the region. In addition, if there are a significant number of pixels with differing labels in a region, these differing labels could be scattered in a speckle-type pattern with a rough texture or they could be clumped together exhibiting smooth texture. In order to characterize the roughness of the labeling texture in a region we will define measures of what we call "speckle."

One clutter-type measure to characterize smoothness (or homogeneity) of classification is easily computed from the typical confusion matrices generated to measure classification performance. The clutter measure, $K_{Clutter}$, is defined as

$$K_{Clutter} = \frac{1}{N_{GTsites}} \sum_i \max\left\{\Upsilon_j^{(i)}, j \in \Gamma\right\}, \quad i = 1, N_{GTreg} \tag{12}$$

where $\Upsilon_j^{(i)}$ is the number of sites assigned to the jth label in the ith ground truth (GT) region, N_{GTreg} is the number of GT regions being evaluated and $N_{GTsites} = \sum_i N_{GTsites}^{(i)}$. The quantity $N_{Gtruth}^{(i)}$ is the number of sites contained in the ith GT region and $i = 1, N_{GTreg}$.

As an indication of the roughness of the labeling in a homogenous region, we consider a measure of speckle that counts the number of neighborhood labels that are different from each site in a region. The measure $K_{Speckle}^{(n=2)}$ is defined with respect to a very simple neighborhood $G_s^{(n=2)}$ consisting of two neighbors - the neighbor to the left and the neighbor above – of a site:

$$K_{Speckle}^{(n=2)} = \sum_{s \in S_{GT}} \sum_{r \in G_s^{(n=2)}} \left(1 - \delta_{x_s, x_r}\right) \tag{13}$$

where S_{GT} is the set of ground truth sites.

A second speckle measure $K_{Speckle}^{(n=4)}$ can also be defined with respect to a more symmetric

neighborhood $G_s^{(n=4)}$ consisting of four neighbors - the neighbors to the left and right, as well as

the neighbors above and below – of a site:

$$K_{Speckle}^{(n=4)} = \sum_{s \in S_{GT}} \sum_{r \in G_s^{(n=4)}} \left(1 - \delta_{x_s, x_r}\right) \tag{14}$$

V. RESULTS

Figure 3 and Table IV show the results of the ISODATA clustering method applied to the image

data for six classes. The commission errors for Trees, Grass, and Asphalt are 22.48%, 1.32%,

and 68.66%, respectively. Although ISODATA provides a fairly good partitioning of the scene,

it is overly sensitive to minor vegetative differences in the forested areas, denoted by "D1-D6" in

the ground-truth overlay (Figure 2) as well as by shade. This is apparent with the high

commission errors for Trees and Asphalt. Omission errors also indicate the same problem. In

particular, omission errors show that another label associated with A1 (Asphalt) is prevalent

throughout subregions of D1-D6 that can be attributed to shade. Specifically, 27% (1015 out of

3764) of the D1-D6 pixels are labeled with the same cluster associated with Asphalt. Also

15.7% (559 out of 3559) of the G1-G4 pixels are labeled with the same cluster as associated with

Trees. Furthermore, the clustering does not provide a smooth and spatially consistent mapping

of regions, and exhibits a significant amount of speckling. Measures of smoothness exhibited the

values $K_{Clutter} = 0.22$, $K_{Speckle}^{(n=2)} = 1641$, and $K_{Speckle}^{(n=4)} = 3359$. These measures will be compared to

those of the Gibbs runs at the end of this section (Tables XIII and XIV).

Using observations of spectral angle distances between the various pairs of materials we set $K_{Thresh}^{(1)} = 11.0$, which remained constant throughout those partition trials involving the angle measure. Using observations of Euclidean distances between the various pairs of materials, we set $K_{Thresh}^{(2)} = 100.0$, which also remained constant throughout those partition trials involving that measure. For both measures and most of the pair-wise signature combinations, there is an obvious gap between similar materials and different materials, except for combinations involving a vegetation signature. Unfortunately, vegetation has a fairly large variance, which makes the threshold boundary a bit fuzzy. During these trials, we observed that the exact value of these thresholds did not seem to be very critical.

Our initial thought was that the values of $K_{Thresh}^{(3)}$ and $K_{Thresh}^{(4)}$ could be selected analytically based on the theoretical properties of the KS statistic. However setting the value of $K_{Thresh}^{(3)}$ to any of the commonly used α values (probabilities of false rejection) resulted in the algorithm finding a significant difference between almost everything. Consequently, the algorithm using the classical KS statistic based on standard α values could not generate a partition map containing any of the structure in the scene. Our interpretation is that the KS statistic is very conservative about accepting a hypothesis that two distributions are similar, and that most of the observed intensity differences were statistically significant. After a number empirical runs, we observed that meaningful structures only occurred at KS distances well beyond the usual theoretical values, and we determined empirically that a value of $K_{Thresh}^{(3)} = 0.5$ was suitable for a spatial window size of 5x5.

Figure 4 and Table V shows the results of Trial 1 testing the spectral angle measure in the energy function. This trial was successful in characterizing the structure of many of the natural and cultural features in the scene without much of the undesirable clutter prevalent in the ISODATA scene shown in Figure 3. For example, the algorithm was not overly sensitive to minor shade differences such as in the forested areas, denoted by "D1-D6" in the ground-truth map. However, the algorithm was also insensitive to other differences that we would like to discriminate. For example, there is missing structure within the road intersection, denoted by "V1," and the rooftops of the apartment buildings (asphalt shingles) denoted by "J2" are confused with the surrounding asphalt parking lot "A3." The commission errors for Trees, Grass, and Asphalt are 4.9%, 2.16%, and 33.07%, respectively. Measures of smoothness improved over the ISODATA trial, exhibiting the values $K_{Clutter} = 0.017$, $K_{Speckle}^{(n=2)} = 222$, and $K_{Speckle}^{(n=4)} = 479$.

Figure 5 and Table VI shows the results of Trial 2 using both the spectral angle and Euclidean distance measures in the energy function. A comparison of Figures 4 and Figure 5 shows a clear advantage to using the two measures in the energy function. Many of the cultural structures missed in Trial 1 are detected in this trial without introducing undesirable vegetation clutter. For example, the structure in "V1" is now evident and so are the rooftops in "J2." The commission errors for Trees, Grass, and Asphalt are 24.08%, 3.88%, and 20.40%, respectively. There has been a significant drop in the commission error of the partition associated with Asphalt compared to the ISODATA trial. Regarding omission error, Table VI shows that the regions D1-D6 are labeled quite homogeneously and are not adversely affected by shade, as was the case

with ISODATA. Specifically, 0.9% (34 out of 3764) of the D1-D6 pixels are labeled with the same cluster associated with Asphalt. Unfortunately, there is a significant problem with the grass region G4, where 818 out of the 832 GT sites were assigned the label associated with Trees. Measures of smoothness have improved over the ISODATA trial, with values

$K_{Clutter} = 0.03$, $K_{Speckle}^{(n=2)} = 395$, and $K_{Speckle}^{(n=4)} = 795$.

Figures 6 and 7 show the results of the single-measure KS trials using a 5x5 spatial window. Figure 6 shows the results of Trial 3, where the classical KS statistic is tested using an empirically-derived value $K_{Thresh}^{(3)} = 0.5$, as mentioned above. Trial 3 was the best performing trial over a number of other similar experimental trials at 5x5 windows (not shown) using different values of $K_{Thresh}^{(3)}$. Nevertheless, the results poorly represent the structure in the scene. Figure 7 shows the results of Trial 4, where the mean-adjusted KS statistic is tested using an analytically-derived value of $K_{Thresh}^{(4)}$, as mentioned above. Trial 4 provides a very reasonable partitioning of the scene, particularly when considering that only one band of imagery is used in the analysis. This measure detects some structure in the building denoted by feature "B1" in the ground-truth map, not found by the other measures, as well as the cars in the parking lot of feature "A1." However, the mean-adjusted KS measure missed a number of cultural features and was unable to distinguish between Grass and Trees. Edge effects are also prevalent throughout the partition map. Not shown are the results of similar trials conducted using a 7x7 spatial window that generally produced inferior results compared to the 5x5 window.

Figure 8 and Table VII shows the results of Trial 5, which combines the spectral angle, Euclidean distance, and mean-adjusted KS statistic measures in the energy function. The

partitioning exhibits more detail than any of the other measures; however, edge effects are prevalent and give the impression of too much clutter, particularly in the cultural regions. Measures of smoothness have degraded compared to the previous Gibbs trials (Trial 1 and Trial 2), exhibiting values of $K_{Clutter} = 0.07$, $K_{Speckle}^{(n=2)} = 665$, and $K_{Speckle}^{(n=4)} = 1377$. This measure might prove to be useful, in the future, if the energy function could be modified to discourage the development of edges. The use of penalty terms in the energy function, perhaps of the form suggested by Geman *et. al.* [9] might be beneficial in removing such artifacts. On the other hand, the tendency for this measure to produce edge artifacts might be in fact be suggesting an alternative use for this energy function as a measure in an edge inducing algorithm -- perhaps a Gibbs-based boundary detection algorithm.

All the above trials for Figures 4-8 were performed using identical control parameters, as listed in Table IIIa, in order to provide some consistency for comparing the different methods. The results indicate that an energy function using the combined spectral angle and Euclidean distance measures is the preferred method, unless an energy function involving the KS measure could somehow be modified to discourage the development of unwanted edges. Based on these results, the subsequent trials (Trials 6-9) were focused on a specific energy function – the one that combines the spectral angle and Euclidean distance measures.

Figure 9 shows the partition map resulting from Trial 6, where the Gibbs algorithm is initialized by the results from the ISODATA algorithm, and Table VIII lists quantitative results. The commission errors for Trees, Grass, and Asphalt are 6.09%, 3.43%, and 20.19%, respectively. So there has been a significant drop in the commission error of the partition associated with

Asphalt, as well as for Trees, compared to the ISODATA trial. Measures of smoothness have improved compared to the ISODATA trial, with values of $K_{Clutter} = 0.0334$, $K_{Speckle}^{(n=2)} = 372$, and $K_{Speckle}^{(n=4)} = 794$. These results show that Trial 6 provides substantial improvement over the results of the standard ISODATA trial (Figure 3 and Table IV).

We now shift attention to studying some additional properties of the algorithm. Given that the combined spectral angle and Euclidean distance measure has shown the most promise, this aspect is fixed for the remaining trials, and we now look at constraining the multigrid process and also the examine the effect of using the more typical compact neighborhood structure in the energy function.

Figure 10 and Table IX show the results of Trial 7. As mentioned in Section IVB, this trial is similar to Trial 2 except that it that it is constrained and it is run at higher initial temperatures during the intermediate stages. A comparison of the results with Figure 5 and Table VI show a better global labeling using the constrained algorithm. The commission errors for Trees, Grass, and Asphalt are .94%, 3.03%, and 21.1%, respectively. The omission errors also improved compared to Trial 2. Specifically, 1.5% (57 out of 3764) of the D1-D6 pixels are labeled with the same cluster associated with Asphalt. Also 0.1% (4 out of 3559) of the G1-G4 pixels are labeled with the some other cluster and the large omission error for G4 in Trial 2 has been reduced to 1.7%. Measures of smoothness for Trial 7 were slightly degraded compared to Trial 2, exhibiting values of $K_{Clutter} = 0.0424$, $K_{Speckle}^{(n=2)} = 536$, and $K_{Speckle}^{(n=4)} = 1093$.

Figure 11 and Table X show the results of Trial 8. This trial was made using the same parameters as Trial 7, except that the typical compact neighborhood is used in the energy function rather than our proposed one. A comparison of these results with Figure 10 and Table IX shows a significant degradation in global labeling. For the trees, in Trial 8, the GT regions D1, D3, and D4 are assigned different labels than D5-D6. For the grasses, in Trial 8, the GT regions for G1/G3, G2, G4, G5, and G6 are assigned different class labels. The commission errors for Trees, Grass, and Asphalt are 21.45%, 34.08%, and 81.32%, respectively. Measures of smoothness for Trial 8 were slightly improved over Trial 7, exhibiting values of $K_{Clutter} = 0.0401$, $K_{Speckle}^{(n=2)} = 467$, and $K_{Speckle}^{(n=4)} = 930$.

Figure 12 and Table XI show the results for Trial 9. This trial was made using the same parameters as Trial 8, except for the addition of one coarser stage in the multigrid process. The addition of this stage was made in an attempt to extend the reach of the compact 5x5 neighborhood by beginning the hierarchical processing at a coarser scale. Visually, the class map seems slightly better than the one for Trial K. However, the global labeling is still significantly degraded from the results achieved in Trial 7 or Trial 2. In fact, the task of making a correspondence between a GT label and a partition label for purposes of establishing commission and omission errors becomes a problem for Grass and Asphalt. Indeed, for purposes of commission error we need to treat LBL03 as both Grass and Asphalt, which is troublesome and makes the usefulness of this measure questionable for this trial. The commission errors for Trees, Grass, and Asphalt are 39.03%, 27.82%, and 80.26%, respectively. Measures of smoothness for Trial 9 were almost identical to Trial 8, exhibiting values of $K_{Clutter} = 0.0483$, $K_{Speckle}^{(n=2)} = 477$, and $K_{Speckle}^{(n=4)} = 924$.

Table XII provides a summary of commission errors for three of the predominant GT categories and for eight of the trials. This table shows that the top two trials in terms of global labeling performance are Trials 6 and 7. These trials both used the combined spectral angle and Euclidean distance measures in the energy function. The global labeling of Trial 6 was helped by a non-random initialization. Trial 7 was helped by the use of the constrained multigrid implementation.

Tables XIII and XIV summarize the homogeneity measures. There is clearly an improvement in homogeneity over the ISODATA algorithm for the Gibbs-based trials using the proposed neighborhood structure shown in Figure 1 (Trials 1, 2, 5, 6, and 7). The trial using a compact 5x5 neighborhood also shows notable improvement in homogeneous labeling (Trial 8).

The far right column of Table XIV lists the ratio of $K^{(n=4)}_{Speckle}$ and $K^{(n=2)}_{Speckle}$ as a means to examine any correspondence between the two measures. This column shows $K^{(n=4)}_{Speckle}$ and $K^{(n=2)}_{Speckle}$ are different by a scale factor of approximately two, which is not surprising.

Finally, the results of Trial 7 (the best performing trial) are shown in Figure 14 as a color class map, with Figure 13 showing a red, green, blue (RGB) true-color composite of the scene. Visually, the color map confirms the results discussed above, showing excellent globally-smooth classification labeling for the majority of the regions. The difficulty that the method sometimes has with narrow and small features is also apparent.

V CONCLUSIONS

A number of energy functions were investigated involving the use of the spectral angle measure, the Euclidean distance, the classical Kolmogorov Smirnov (KS) statistic, the mean-adjusted KS measure, and certain combinations of these measures. The results of this experiment indicate that the energy function using a combination of the spectral angle and Euclidean distance measures will produce excellent results. This energy function is the preferred one in the present context.

The tendency for the KS measure (as a single measure, or in combination with other measures) to produce edge artifacts may actually indicate an alternative use for this energy function as a measure suitable for an edge-inducing algorithm -- perhaps a Gibbs-based boundary detection algorithm. It should be reiterated that the KS measure was calculated at a single wavelength. A current avenue being explored is where the KS statistic is calculated as a voxel neighborhood (two spatial and one spectral).

Overall, the Gibbs-based algorithm generates partitions that accurately represent structured regions in the image with locally smooth labeling. For most of the bigger regions, pixels with the same label represent the same type of phenomenon on the ground, globally across the image, and regions with different labels correspond to different phenomenon. A globally consistent labeling of the partitions was sometimes problematic, particularly for some of the small as well as narrow regions. This problem can be reduced significantly by using (1) non-random initialization and/or (2) the constrained multigrid approach. This is explicitly demonstrated in Trials 6 and 7, respectively. The quantitative measures of local consistency (smoothness) and

global labeling, along with the class maps demonstrate a clear benefit of the Gibbs-based method as compared to the traditional ISODATA clustering algorithm.

The proposed neighborhood structure discussed in IIIA was shown to provide significantly better global labeling than the typical MRF compact neighborhood without any significant degradation in labeling smoothness. A comparison of the results of Trials 8 and 9 with the corresponding trials using our proposed non-compact neighborhood structure explicitly make this point.

The Gibbs-based algorithm can be used as a stand-alone method; however, it is also envisioned to work within an overall framework involving other processes. In particular, the partitioning of hyperspectral imagery is viewed as a mechanism to condition a spectral mixing process as proposed in [12]. Under this framework, the ultimate identification of materials (and the corresponding abundances) in a scene is a consequence of the spectral mixing process, so that a precise global labeling is not required at the partitioning stage.

ACKNOWLEDGEMENTS

REFERENCES

[1] Duda R. and Hart P., *Pattern Classification and Scene Analysis*, John Wiley and Sons, 1973.

[2] Therrien C., *Decision Esimation and Classification – An Introduction to Pattern Recognition and Related Topics*, John Wiley and Sons, 1989.

[3] Jimenez, L.O., Velez-Reyes, M., Rivera-Medina, J.L., Velasquez, H.T., "Unsupervised Decision Fusion for Hyperspectral Data," Submitted to the *Proceedings of SPIE- International Symposium on Remote Sensing,* Toulouse, France, September 2001.

[4] Reyes M., Jimenez, Linares D., Velazquez H., "Comparison of matrix factorization algorithms for band selection in hyperspectral imagery," *Proceeding of SPIE Proceedings of the SPIE International Symposium on Aerospace/Defense Sensing, Simulation, and Controls (Aerosense),* Orlando FL., March 2000.

[5] Green A., Berman M., Switzer P., and Craig M., "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 26, January 1988.

[6] Rivera-Medina, J.L., Jimenez, L.O., Velez-Reyes, M., Ramirez-Velez, M. D.,

"Unsupervised Classification of High Spatial and Spectral Resolution Data based on the Unsupervised ECHO Classifier," Proceedings of IGARSS, 2000, Honolulu, Hawaii.

[7] Winkler G., *Image Analysis, Random Fields and Dynamic Monte Carlo Methods – A Mathematical Introduction*, Applications of Mathematics, Springer-Verlag, 1995.

[8] Besag, J., "Towards Bayesian image analysis," *Supplement to Journal of Applied Statistics*, Vol. 20, Nos. 5/6, 1993.

[9] Geman S. and Geman D., Graffigne C., and Dong P., "Boundary detection by constrained optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 7, July 1990.

[10] Geman S. and Geman D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, November 1984.

[11] Gidas B., "A Renormalization Group Approach to Image Processing Problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, No. 2, February 1989.

[12] Rand R. and Keenan D., "A spectral mixture process conditioned by Gibbs-based partitioning," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, No. 7, July 2001.

[13] Aarts E. and Korst, *Simulated Annealing and Bolzmann Machines - A Stochastic Approach to Combinatorial Optimization and Neural Computing*, Interscience Series in Discrete Mathematics and Optimization, John Wiley and Sons, 1989, Reprint 1990.

BIOGRAPHY

Robert S. Rand received a B.S. degree in Physics from the University of Massachusetts at Lowell, a Masters of Engineering Physics degree from the University of Virginia, Charlottesville VA, and his Ph.D. in Engineering Physics from the University of Virginia. He works at the U.S. Department of the Army Engineer Research and Development Center (ERDC), Corps. of Engineers, Topographic Engineering Center, Alexandria, VA. His research interests are in enhancing hyperspectral processing techniques through spectral mixture analysis, image segmentation, neural networks, Markov-Random Field (MRF) models, and Independent Component Analysis.

Daniel M. Keenan received his Ph.D. from the University of Chicago, and is Associate Professor of Statistics in the Department of Statistics at the University of Virginia, Charlottesville, VA. His research interests are in Applied Mathematics, in particular, Pattern Theory, Bayesian image analysis, and modeling in medicine.

Authors' Email:

R. Rand   - Robert.S.Rand@erdc.usace.army.mil
D. Keenan - dmk7b@virginia.edu

Tables

Table I. Description of the ground truth (GT) information, as displayed in Figure 2.

GT Label	Description of the Area
A1	Asphalt parking lot, newly paved.
A2-A3	Asphalt parking lot. A3a and A3b opposite sides of apartment complex, different ages.
A4-A5	Asphalt road
B1	Rooftop of a department store, light-gray asphalt shingles, skylights, yellow gas pipes.
B2a,b	School, rooftop mostly coated bubbly light-gray material (B2b), top edge is gravel (B2a).
C1	Concrete pavement to loading dock, some tire marks.
C2a,b	Concrete parking lot surrounding small bank building. C2b is brighter than C2a.
D1	Trees: Black Willow, Texas Sugar Berry, Dogwood, Texas oak, Elbon bush, Green Bryer.
D3	Trees
D4	Trees: Texas Oak, Burmelia, and Texas Hackberry.
D5	Trees: Red Bud, Texas Oak, and Red Bud.
D6a,b,c	Trees - Deciduous holly, Bois D'Arc, Osage Orange, Texas Oak, and Bow Wood
D7	Scattered junipers throughout the grassy areas.
E1	Church and school with metal rooftop.
E2-E6	Corrugated steel buildings.
G1-G6	Tall wild grass. G5 and G6 near drainage and healthier.
H1	Healthy maintained grass near bank building.
J1	Private residence, stand-alone.
J2	Apartment buildings with asphalt-shingled rooftops.
J3	Church with dark brown asphalt-shingled rooftop.
L1	Shade of Apartment building.
M1	Tennis court, concrete with blue/gray coating, and lots of dirt.
N1	Playground area with exposed soil.
O1	Encircled area is residential with moderately-priced homes and concrete driveways.
R1	Light linear pattern is exposed bedrock.
R2-R3	Exposed soil.
V1	Asphalt road intersection with exposed soil along the shoulders.
W1	Playground - grass with exposed soil, gravel, and some asphalt.

Table II. Descriptive summary of the trials.

Trial	Disparity measure(s)	Description
1	Spectral angle	Unconstrained, 5 stages
2	Spectral angle, Euclidean distance	Unconstrained, 5 stages
3	Classical KS statistic (5x5 window)	Unconstrained, 5 stages
4	Mean-adjusted KS statistic (5x5)	Unconstrained, 5 stages
5	Spectral angle, Euclidean distance, mean-adjusted KS statistic	Unconstrained, 5 stages
6	Spectral angle, Euclidean distance	Initialized by ISODATA
7	Spectral angle, Euclidean distance	Constrained, 5 stages
8	Spectral angle, Euclidean distance (5X5 compact neighborhood)	Constrained, 6 stages
9	Spectral angle, Euclidean distance (5X5 compact neighborhood)	Constrained, 6 stages

Table III. Control parameters for the trials. The sequence of numbers in the "Dist of neighbors" column corresponds to the distance from the center pixel (the "site") to the perimeter, for the surrounding near neighbors, the intermediate neighbors, and the far neighbors. The sequence of numbers in the "# of neighbors" column corresponds to the number of near neighbors, intermediate neighbors, and far neighbors in the neighborhood system.

Table IIIa. Control parameters for Trials 1, 2, 3, 4, 5

σ	C	# iterations	Dist. of neighbors	# of neighbors
16	9.0	50,000	16,64,128	8, 8, 8
8	3.0	2,000	8,32,64	8, 8, 8
4	3.0	2,000	4,16,32	8, 8, 8
2	2.0	700	2,8,16	8, 8, 8
1	1.0	350	1,4, 8	8, 8, 8

Table IIIb. Control parameters for Trial 7.

σ	C	# iterations	Dist. of neighbors	# of neighbors
16	9.0	50,000	16,64,128	8, 8, 8
8	6.0	10,000	8,32,64	8, 8, 8
4	3.0	2,000	4,16,32	8, 8, 8
2	3.0	2,000	2,8,16	8, 8, 8
1	1.0	350	1,4, 8	8, 8, 8

Table IIIc. Control parameters for Trial 8.

σ	C	# iterations	Dist. of neighbors	# of neighbors
16	9.0	50,000	compact	24
8	6.0	10,000	compact	24
4	3.0	2,000	compact	24
2	3.0	2,000	compact	24
1	1.0	350	compact	24

Table IIId. Control parameters for Trial 9.

σ	C	# iterations	Dist. of neighbors	# of neighbors
32	9.0	50,000	compact	24
16	9.0	50,000	compact	24
8	6.0	10,000	compact	24
4	3.0	2,000	compact	24
2	3.0	2,000	compact	24
1	1.0	350	compact	24

Table IV. Confusion matrix of results for the ISODATA algorithm. Labels in the rows are the GT labels, and labels in the columns are the partitioned labels. As an example of how to read the table, the region of ground truth extracted for C1 consisted of 120 pixels. Of the 120 pixels, 13 pixels were assigned to cluster label ISO-04, and 107 pixels were assigned to cluster label ISO-05.

Labels	ISO-1	ISO-02	ISO-03	ISO-04	ISO-05	ISO-06	NULL
A1	0	0	518	0	0	0	0
B1	0	0	0	278	0	0	0
C1	0	0	0	13	107	0	0
C2a	0	0	0	0	13	0	0
C2b	0	0	0	0	11	0	0
B2a	0	0	0	2	15	0	0
B2b	0	0	0	12	110	0	0
M1	0	0	0	43	0	0	0
J3	0	0	33	44	0	0	0
A2	0	4	0	51	0	0	0
A3a	0	0	0	16	0	0	0
A3b	0	0	0	12	0	0	0
E1	0	0	0	0	29	0	0
E3	0	0	0	11	13	0	0
E4	0	0	0	0	15	0	0
E5	0	0	0	0	71	0	0
E6	0	0	0	22	8	0	0
J2	0	0	84	0	0	0	0
N1	0	0	0	5	88	0	0
R3	0	0	0	0	56	0	0
D1	203	0	68	0	0	1	0
D3	231	8	138	0	0	9	0
D4	859	11	377	0	0	98	0
D5	262	6	108	0	0	54	0
D6a	279	4	43	0	0	233	0
D6b	346	5	261	0	0	41	0
D6c	96	2	20	0	0	1	0
G1	0	504	0	0	0	8	0
G2	87	969	0	0	0	0	0
G3	13	1143	3	0	0	0	0
H1	0	0	0	0	0	33	0
W1	0	1	0	65	0	0	0
G4	459	373	0	0	0	0	0
G5	101	0	0	0	0	71	0
G6	0	10	0	0	0	176	0
Total GT	2936	3040	1653	574	536	725	0
Total non-GT	27775	55909	29245	26055	10402	21150	0

Table V. Confusion matrix of results for Trial 1 - the spectral angle disparity measure. Labels in the rows are the GT labels, and labels in the columns are the partitioned labels.

Labels	LBL-01	LBL-02	LBL-03	LBL-04	LBL-05	LBL-06	NULL
A1	0	0	518	0	0	0	0
B1	0	0	0	277	1	0	0
C1	0	0	120	0	0	0	0
C2a	0	0	7	0	0	0	0
C2b	0	0	11	0	0	0	0
B2a	0	0	0	17	0	0	0
B2b	0	0	0	122	0	0	0
M1	0	0	0	2	41	0	0
J3	0	0	0	0	0	77	0
A2	0	0	55	0	0	0	0
A3a	0	0	16	0	0	0	0
A3b	0	0	12	0	0	0	0
E1	0	0	0	29	0	0	0
E3	0	0	0	0	0	24	0
E4	0	0	0	15	0	0	0
E5	0	0	0	0	71	0	0
E6	0	0	0	0	2	28	0
J2	0	0	84	0	0	0	0
N1	0	1	0	92	0	0	0
R3	0	0	56	0	0	0	0
D1	272	0	0	0	0	0	0
D3	348	27	0	2	6	3	0
D4	1308	26	7	0	2	2	0
D5	412	6	1	6	0	5	0
D6a	551	3	2	2	0	1	0
D6b	620	10	7	10	1	5	0
D6c	117	0	2	0	0	0	0
G1	0	0	0	512	0	0	0
G2	0	1054	0	1	0	1	0
G3	0	1158	0	1	0	0	0
H1	0	33	0	0	0	0	0
W1	0	66	0	0	0	0	0
G4	15	817	0	0	0	0	0
G5	172	0	0	0	0	0	0
G6	0	185	0	0	1	0	0
Total GT	3815	3386	898	1088	125	146	0
Total non-GT	41033	65311	30630	13467	9636	5009	5456

Table VI. Confusion matrix of results for Trial 2 – the combined spectral angle/Euclidean distance disparity measure. Labels in the rows are the GT labels, and labels in the columns are the partitioned labels.

Labels	LBL-01	LBL-02	LBL-03	LBL-04	LBL-05	LBL-06	NULL
A1	0	0	518	0	0	0	0
B1	0	0	0	278	0	0	0
C1	0	0	0	0	120	0	0
C2a	6	0	0	0	0	1	6
C2b	11	0	0	0	0	0	0
B2a	0	0	0	0	17	0	0
B2b	0	0	0	0	122	0	0
M1	0	0	41	0	2	0	0
J3	0	0	77	0	0	0	0
A2	0	0	55	0	0	0	0
A3a	0	0	16	0	0	0	0
A3b	0	0	12	0	0	0	0
E1	9	1	0	0	19	0	0
E3	0	0	0	7	0	17	0
E4	0	15	0	0	0	0	0
E5	29	0	0	0	0	42	0
E6	2	0	0	24	0	4	0
J2	0	0	0	0	84	0	0
N1	0	1	0	0	92	0	0
R3	56	0	0	0	0	0	0
D1	272	0	0	0	0	0	0
D3	345	25	0	4	12	0	0
D4	1278	32	8	14	6	7	0
D5	394	15	13	4	2	2	0
D6a	527	14	5	4	4	5	0
D6b	601	10	8	12	11	11	0
D6c	117	0	0	0	0	2	0
G1	0	512	0	0	0	0	0
G2	0	1054	0	1	0	1	0
G3	0	1159	0	0	0	0	0
H1	0	33	0	0	0	0	0
W1	0	66	0	0	0	0	0
G4	818	0	0	0	14	0	0
G5	0	0	0	0	172	0	0
G6	185	1	0	0	0	0	0
Total GT	4655	2939	755	352	680	98	6
Total non-GT	41394	64214	20247	15298	14026	9904	5453

Table VII. Confusion matrix of results for Trial 5 - the combined spectral angle, Euclidean distance, and mean-adjusted KS disparity measures. Labels in the rows are the GT labels, and labels in the columns are the partitioned labels.

Labels	LBL-01	LBL-02	LBL-03	LBL-04	LBL-05	LBL-06	NULL
A1	0	2	513	2	1	0	0
B1	0	9	0	260	1	8	0
C1	0	0	0	2	118	0	0
C2a	0	0	3	1	0	3	0
C2b	0	10	1	0	0	0	0
B2a	3	2	2	3	2	5	0
B2b	3	4	1	0	0	114	0
M1	4	0	0	0	39	0	0
J3	16	14	10	10	14	13	0
A2	0	0	0	2	0	53	0
A3a	0	0	0	0	0	16	0
A3b	0	8	0	0	0	4	0
E1	11	3	2	10	2	1	0
E3	0	0	14	5	0	5	0
E4	3	0	0	0	8	4	0
E5	17	17	13	8	16	0	0
E6	26	0	0	1	3	0	0
J2	61	6	5	8	4	0	0
N1	3	89	1	0	0	0	0
R3	0	53	0	2	0	1	0
D1	0	0	0	0	0	272	0
D3	344	25	0	4	5	8	0
D4	20	9	1275	10	10	21	0
D5	9	11	4	2	8	396	0
D6a	521	4	12	1	5	16	0
D6b	597	5	17	10	10	14	0
D6c	115	0	0	2	1	1	0
G1	1	12	494	0	4	1	0
G2	1	1046	1	6	0	2	0
G3	61	0	1097	0	0	1	0
H1	1	0	32	0	0	0	0
W1	0	0	66	0	0	0	0
G4	0	824	8	0	0	0	0
G5	0	1	166	0	0	5	0
G6	4	83	0	1	0	98	0
Total GT	1821	2237	3737	350	251	1062	0
Total non-GT	26702	37175	38199	19193	22217	21544	5512

Table VIII. Confusion matrix of results for Trial 6 - the combined spectral angle and Euclidean distance disparity measures, when initialized by the ISODATA algorithm. Labels in the rows are the GT labels, and labels in the columns are the partitioned labels.

Labels	LBL-01	LBL-02	LBL-03	LBL-04	LBL-05	LBL-06	NULL
A1	0	0	518	0	0	0	0
B1	0	0	0	278	0	0	0
C1	0	0	0	0	120	0	0
C2a	5	2	0	0	6	0	0
C2b	11	0	0	0	0	0	0
B2a	0	0	0	0	17	0	0
B2b	0	0	0	1	121	0	0
M1	0	0	0	41	2	0	0
J3	0	0	0	77	0	0	0
A2	0	0	55	0	0	0	0
A3a	0	0	0	16	0	0	0
A3b	0	0	0	12	0	0	0
E1	19	1	0	9	0	0	0
E3	0	0	24	0	0	0	0
E4	0	0	0	0	0	15	0
E5	0	42	0	0	0	29	0
E6	0	4	0	0	1	25	0
J2	0	0	84	0	0	0	0
N1	0	1	0	0	92	0	0
R3	0	0	0	0	56	0	0
D1	272	0	0	0	0	0	0
D3	343	27	1	0	1	14	0
D4	1284	22	15	7	4	13	0
D5	392	14	3	7	9	5	0
D6a	524	6	11	6	5	7	0
D6b	598	9	7	12	8	19	0
D6c	117	0	0	0	2	0	0
G1	0	512	0	0	0	0	0
G2	0	1054	0	0	0	2	0
G3	0	1159	0	0	0	0	0
H1	0	0	0	0	33	0	0
W1	0	66	0	0	0	0	0
G4	22	810	0	0	0	0	0
G5	172	0	0	0	0	0	0
G6	0	3	0	0	0	183	0
Total GT	3759	3732	718	466	477	312	0
Total non-GT	36442	71392	18126	15793	13671	15112	0

Table IX. Confusion matrix of results for Trial 7 – Constrained 5-stage multigrid partitioning. Labels in the rows are the GT labels, and labels in the columns are the partitioned labels.

Labels	LBL-01	LBL-02	LBL-03	LBL-04	LBL-05	LBL-06	NULL
A1	0	0	518	0	0	0	0
B1	0	0	0	278	0	0	0
C1	0	0	0	0	0	120	0
C2a	0	0	0	4	3	0	6
C2b	0	0	0	2	9	0	0
B2a	0	0	0	0	0	17	0
B2b	0	0	0	1	0	121	0
M1	1	0	0	0	42	0	0
J3	0	0	0	0	76	1	0
A2	0	0	55	0	0	0	0
A3a	0	0	6	10	0	0	0
A3b	0	0	0	12	0	0	0
E1	0	0	0	2	9	18	0
E3	12	0	6	0	0	6	0
E4	5	0	0	0	0	10	0
E5	1	29	0	0	41	0	0
E6	0	1	1	0	28	0	0
J2	0	0	84	0	0	0	0
N1	0	7	6	0	0	80	0
R3	0	0	0	0	56	0	0
D1	268	0	1	2	0	1	0
D3	340	27	3	7	7	2	0
D4	1264	26	22	4	18	11	0
D5	391	12	9	3	12	3	0
D6a	522	3	14	14	3	3	0
D6b	588	12	8	13	22	10	0
D6c	117	2	0	0	0	0	0
G1	0	512	0	0	0	0	0
G2	1	1053	0	0	1	1	0
G3	0	1159	0	0	0	0	0
H1	0	33	0	0	0	0	0
W1	0	59	0	0	0	7	0
G4	1	818	0	0	13	0	0
G5	0	0	0	0	172	0	0
G6	12	170	1	0	3	0	0
Total GT	3523	3923	734	352	515	411	6
Total non-GT	36776	69343	19405	12768	12078	14725	5441

Table X. Confusion matrix of results for Trial 8 - Constrained 5 stage multigrid partitioning using a compact 5x5 neighborhood. Labels in the rows are the GT labels; and labels in the columns are the partitioned labels.

Labels	LBL-01	LBL-02	LBL-03	LBL-04	LBL-05	LBL-06	NULL
A1	0	0	518	0	0	0	0
B1	278	0	0	0	0	0	0
C1	0	0	0	0	119	1	0
C2a	0	0	0	0	5	0	8
C2b	0	0	0	0	11	0	0
B2a	0	0	0	17	0	0	0
B2b	0	0	0	122	0	0	0
M1	42	0	1	0	0	0	0
J3	0	77	0	0	0	0	0
A2	54	1	0	0	0	0	0
A3a	0	8	0	0	8	0	0
A3b	0	0	0	0	11	1	0
E1	0	1	0	9	0	19	0
E3	0	16	0	8	0	0	0
E4	0	0	0	15	0	0	0
E5	41	0	0	0	30	0	0
E6	28	1	0	0	0	1	0
J2	0	84	0	0	0	0	0
N1	0	0	92	0	1	0	0
R3	0	0	56	0	0	0	0
D1	0	270	1	0	0	1	0
D3	1	30	341	7	2	5	0
D4	4	14	19	1272	22	14	0
D5	394	13	5	6	6	6	0
D6a	518	18	2	4	6	11	0
D6b	592	13	7	8	14	19	0
D6c	117	0	0	0	0	2	0
G1	0	1	508	1	0	0	2
G2	0	1055	0	1	0	0	0
G3	1	0	1158	0	0	0	0
H1	0	0	0	0	0	33	0
W1	0	0	65	1	0	0	0
G4	0	0	0	821	11	0	0
G5	0	0	0	0	172	0	0
G6	0	0	0	40	146	0	0
Total GT	2070	1602	2773	2332	564	113	10
Total non-GT	25251	29977	47107	26408	17295	17372	7126

Table XI. Confusion matrix of results for Trial 9 - Constrained 6 stage multigrid partitioning using a compact 5x5 neighborhood. Labels in the rows are the GT labels, and labels in the columns are the partitioned labels.

Labels	LBL-01	LBL-02	LBL-03	LBL-04	LBL-05	LBL-06	NULL
A1	0	0	518	0	0	0	0
B1	0	0	0	0	278	0	0
C1	120	0	0	0	0	0	0
C2a	0	0	5	0	0	0	8
C2b	0	0	11	0	0	0	0
B2a	17	0	0	0	0	0	0
B2b	122	0	0	0	0	0	0
M1	0	0	0	1	0	42	0
J3	0	0	0	0	77	0	0
A2	54	1	0	0	0	0	0
A3a	0	0	8	8	0	0	0
A3b	0	0	12	0	0	0	0
E1	1	6	0	7	3	12	0
E3	0	16	0	8	0	0	0
E4	0	0	0	15	0	0	0
E5	42	0	0	0	29	0	0
E6	0	28	1	0	1	0	0
J2	0	0	0	83	0	1	0
N1	0	93	0	0	0	0	0
R3	56	0	0	0	0	0	0
D1	1	270	0	0	0	1	0
D3	81	28	7	265	5	0	0
D4	4	8	28	18	1269	18	0
D5	5	12	396	9	0	8	0
D6a	10	3	514	5	17	10	0
D6b	12	13	600	12	6	10	0
D6c	2	0	117	0	0	0	0
G1	0	1	509	0	0	0	2
G2	0	1055	0	1	0	0	0
G3	2	0	0	0	0	1157	0
H1	0	0	0	33	0	0	0
W1	0	0	0	66	0	0	0
G4	0	12	0	0	818	2	0
G5	0	172	0	0	0	0	0
G6	0	0	0	147	39	0	0
Total GT	529	1718	2726	678	2542	1261	10
Total non-GT	18699	36006	32998	25531	29256	20920	7126

Table XII. Errors of commission for eight trials Each column lists the percentages of labels assigned to a

partition belonging to a GT category other than the one listed at the top of a column.

Commission Errors			
Trial	Trees	Grass	Asphalt
ISODATA	22.48%	1.32%	68.66%
Trial 1	4.90%	2.16%	33.07%
Trial 2	24.08%	3.88%	20.40%
Trial 5	11.81%	12.11%	86.27%
Trial 6	6.09%	3.43%	20.19%
Trial 7	0.94%	3.03%	21.12%
Trial 8	21.45%	34.08%	81.32%
Trial 9	39.03%	27.82%	80.26%

Table XIII. Measures of clutter. $K_{Clutter}$ are listed for eight of the trials.

Trial	$K_{Clutter}$
ISODATA	0.2263
Trial 1	0.0170
Trial 2	0.0320
Trial 5	0.0720
Trial 6	0.0334
Trial 7	0.0424
Trial 8	0.0401
Trial 9	0.0483

Table XIV. Measures of speckle. Two smooth measures of speckle, $K_{Speckle}^{(n=2)}$ and $K_{Speckle}^{(n=4)}$, are listed for eight of

the trials, along with the ratio of the two measures.

Trial	$K_{Speckle}^{(n=2)}$	$K_{Speckle}^{(n=4)}$	$K_{Speckle}^{(n=4)} / K_{Speckle}^{(n=2)}$
ISODATA	1641	3359	2.047
Trial 1	222	479	2.158
Trial 2	395	795	2.013
Trial 5	665	1377	2.071
Trial 6	372	794	2.134
Trial 7	536	1093	2.039
Trial 8	467	930	1.991
Trial 9	477	924	1.937

Figures
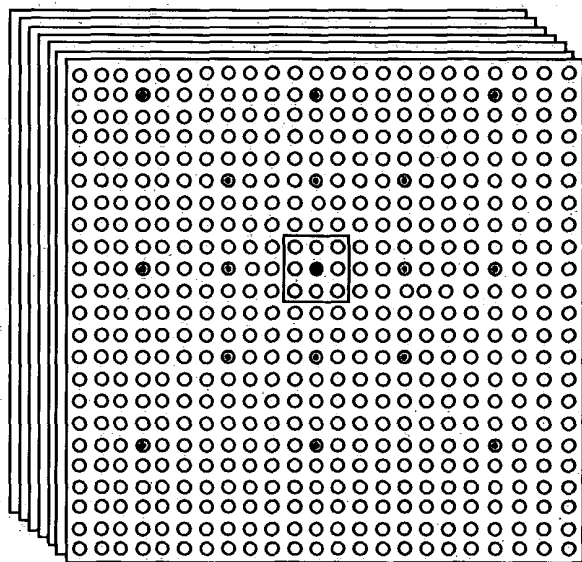
Figure 1 A neighborhood system G_s at site s.

Figures
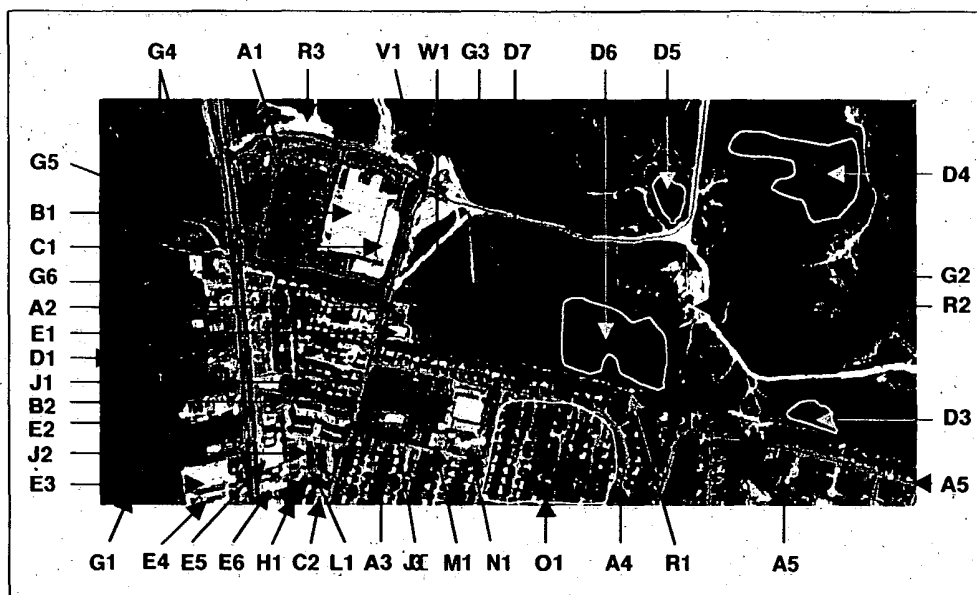
Figure 2. HYDICE Band 49 with Ground truth overlay. FIG. 1

Figure 3. ISODATA cluster map. FIG. 2

Figure 4. Trial 1 - Unconstrained method using spectral angle measure.

Figure 5. Trial 2 - Unconstrained method using spectral angle and Euclidean distance measures.
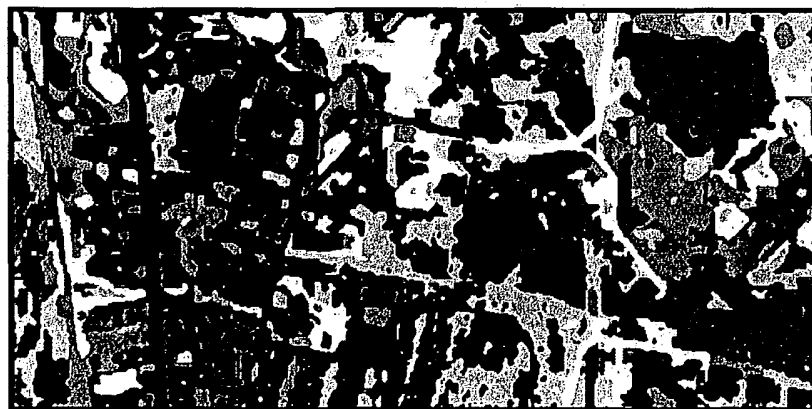
Figure 6. Trial 3 – Unconstrained method using classical KS measure

Figure 7. Trial 4 – Unconstrained method using mean adjusted KS measure.

Figure 8. Trial 5 – Unconstrained method using spectral angle, Euclidean distance, and adjusted KS measures.

Figure 9. Trial 6 – Unconstrained method initialized by ISODATA using spectral angle and Euclidean distance measures.

Figure 10. Trial 7 – Constrained 5-stage method using two disparity measures.

Figure 11. Trial 8 - Constrained 5-stage method using two disparity measures but a compact neighborhood.
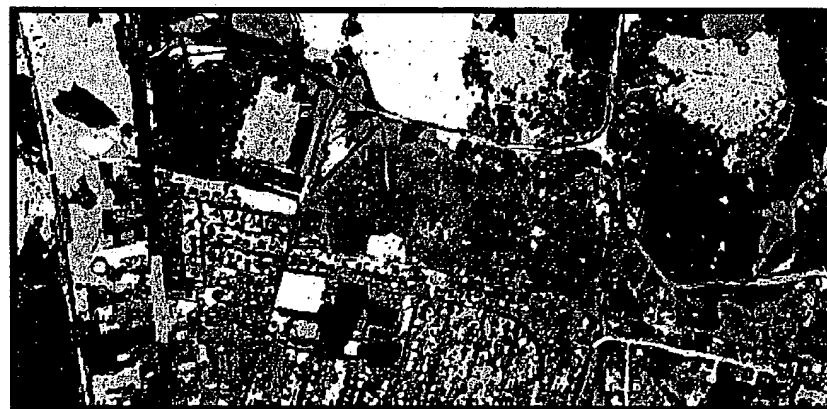
Figure 12. Trial 9 - Constrained 6-stage method using two disparity measures but a compact neighborhood.

Figure 13. Color Composite of HYDICE Bands 49, 35, 18 (RGB).
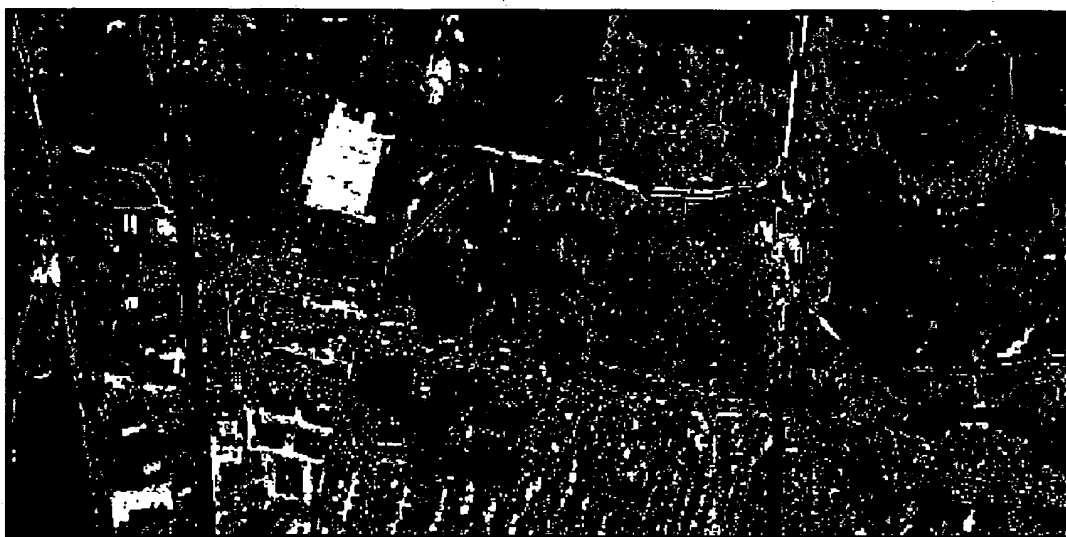
Figure 14. Color Class Map of Trial 7 Results Constrained 5-stage method using two disparity measures.

APPENDIX F

Spectral/spatial annealing of hyperspectral imagery initialized by a supervised classification method

Robert S. Rand

U.S. Army Engineer Research and Development Center, Alexandria, VA 22315-3864

ABSTRACT

A simulated annealing method of partitioning hyperspectral imagery, initialized by a supervised classification method, is investigated to provide spatially smooth class labeling for terrain mapping applications. The method is used to obtain an estimate of the mode a Gibbs distribution defined over a symmetric spatial neighborhood system that is based on an energy function characterizing spectral disparities in Euclidean distance and spectral angle. Experiments are conducted on a 210-band HYDICE scene that contains a diverse range of terrain features and that is supported with ground truth. Both visual and quantitative results demonstrate a clear benefit of this method as compared to spectral-only supervised classification or unsupervised annealing that has been initialized randomly.

Keywords: Gibbs distribution, Bayesian, Markov Random Field, supervised classification, Euclidean distance, Spectral Angle, clustering, hyperspectral imagery.

1. INTRODUCTION

A Bayesian spectral/spatial approach to segmenting hyperspectral imagery into homogeneous regions is investigated where spatial consistency is imposed on the spectral content of sites in the scene. The intention is to produce a spatially-smooth labeling of regions in a scene with minimal loss of significant information and scene content. For example, we desire to eliminate as much insignificant clutter as possible (such as the mottled tree patterns in forested regions) while retaining the small, but significant, terrain features (i.e. a small rock outcrop, or building, etc.).

In prior work,[1,2] we investigated a Bayesian Gibbs-based unsupervised approach to partitioning hyperspectral imagery that is initialized randomly. Various energy functions using different disparity measures were investigated, the merit of using a non traditional extended neighborhood system and a Renormalization Group (RB) based multigrid implementation to improve global labeling was demonstrated, and an enhanced method of spectral mixture analysis (SMA) making use of such a partitioning process was proposed that is appropriate for terrain mapping when the imagery involved is characterized by moderate-to-high scene complexity. Whereas, the results have been very encouraging as compared to other existing unsupervised methods, such as KMEANS and ISODATA,[3,4] concerns remain about obtaining a truly global labeling solution for practical computing times, and there has been a nagging problem with the method having a hard time extracting small narrow terrain features, such as roads. However, as we have repeatedly pointed out, we only examined its behavior for the case of random initialization, which is really a "worst-case" scenario.

The current effort extends the development of the unsupervised spectral/spatial approach by initializing a partitioning algorithm with another Bayesian-based classification process. By changing the assumptions governing the class-conditional and prior probability distributions we recover a traditional spectral-only supervised process that can provide spectrally-optimal initial estimates for the spectral/spatial partitioning process. Initializing the partitioning algorithm in this manner effectively makes the method a supervised classifier.

The prior and current work has its foundations based on the pioneering efforts of others in the Bayesian community. A general mathematical development of the underlying theory, applying the Bayesian framework in the context of image processing, is well established.[5] Using such a framework, the properties of a Markov Random Field (MRF) have been

applied to develop a Gibbs-based partitioning algorithm for gray-scale imagery with two spatial coordinates, and the Gibbs sampler and Metropolis methods have been used as ways to compute maximum a posteriori (MAP) estimates.[5,6,7]

2. DESCRIPTION OF ALGORITHM

2.1 Bayesian Framework

A Bayesian approach provides our framework to construct an algorithm for the classification labeling of homogeneous regions in hyperspectral imagery. Consider the combined process $X = \left(X^P, X^\lambda \right)$ where X^P is a discrete labeling process that provides class labels that identify similar regions in a scene, and X^λ is a spectral modeling process for the underlying spectra of materials in a scene of hyperspectral data. What is actually observed is G, the observed hyperspectral image cube. For X^λ and G there are both spatial and spectral coordinates.

The algorithm is generalized to be multigrid, and thus we define a collection of lattices that provides a finite sequence of coarser and coarser grids: $S_P^{(\sigma_1)} \to S_P^{(\sigma_2)} \to \ldots \to S_P^{(\sigma_{N_{Seq}})}$, where

$$S_P^{(\sigma_n)} = \left\{ (i\sigma_n +1, j\sigma_n +1): \ 0 \le i \le \tfrac{N_{Rows}-1}{\sigma_n}, \ 0 \le j \le \tfrac{N_{Cols}-1}{\sigma_n} \right\} \tag{1}$$

N_{Seq} is the number of stages in the sequence, σ_n is a parameter called the "grid resolution" which determines the spatial sampling of the algorithm and the spatial resolution (coarseness) of the labeling process at stage n ($n = 1, \ldots, N_{Seq}$) of the multigrid processing, N_{Rows} is the number of rows in an image, and N_{Cols} is the number of columns. $S_P^{(\sigma_n)}$ is the set of sites that will be used in the partition process at a resolution σ_n. Our grid structure is a special case of the general course-to-fine grid structure used in the RGA approach to image processing,[8] where the coarsest level contains all the large-scale features and each finer level contains correspondingly smaller-scale features. The term "site" refers to a generic element of a lattice, as specified in (1). At full-resolution a site corresponds to a pixel, but at a coarser scale it corresponds to a square block of pixels. The value of σ_n is defined to be the number of pixels per side of each block of pixels. The finest resolution is $\sigma_1 = 1$. We denote the full-resolution image pixel lattice as S_I, i.e., the resolution at which the observations are obtained, where $S_I = S_P^{\sigma_1}$.

From a Bayesian perspective, the classification problem can be viewed as maximizing the posterior probability distribution $P\left(X^P \mid X^\lambda \right)$, which can be written

$$\Pr\left(X^P \mid X^\lambda \right) \propto \Pr\left(X^\lambda \mid X^P \right) \Pr\left(X^P \right) \tag{2}$$

where $\Pr\left(X^\lambda \mid X^P \right)$ is the class conditional distribution and $\Pr\left(X^P \right)$ is the prior distribution. X^P is a discrete labeling process which associates a label with each site. The realizations of X_s^P are $x_s^P \in \Gamma$, where $\Gamma = \{1, 2, \ldots, N_{Labels}\}$ is the set of classification labels. N_{Labels} is the total number of labels assigned to the process. The configuration space for this labeling process X^P is then determined by N_{Labels} and σ_n (the label resolution), and it is denoted by $\Omega_P^{(\sigma_n)} = \left\{ x^P = (x_s^P, \ s \in S_P^{(\sigma_n)}) \mid x_s^P \in \left\{ 1, \ldots, N_{Labels} \right\} \right\}$.

2.2 The Partitioning Algorithm

We now make the assumption of uncorrupted data such that $X^\lambda = G$ and of singular correspondence between data assignments and class membership $P(G = g \mid X^P = x^P) = 1$, where g are the realizations of G and x^P are the realizations of X^P. Therefore, (2) simplifies to $\Pr(X^P \mid X^\lambda) \propto \Pr(X^P)$, and the classification problem reduces to maximizing $\Pr(X^P)$. Under these assumptions we refer to the labeling process as the "Partition Process."

Furthermore, we treat X^P as a Markov Random Field (MRF) with respect to a graph $\{S_P^{(\sigma_n)}, G^{(\sigma_n)}\}$, where $G^{(\sigma_n)}$ is the entire neighborhood structure for the partition process at resolution σ_n, and G_s is the specific neighborhood at site s for the partition process at resolution σ_n. For any S that is one of $S_P^{(\sigma_n)}$: $S = \{s_1, \ldots, s_M\}$, where M is the number of sites on the grid, an associated neighborhood system $G^{(\sigma_n)} = \{G_s, s \in S\}$ is specified which is a set of lattice graph structures of neighbors, such that $s \notin G_s$, and $s \in G_r$ if and only if $r \in G_s$.

In constructing an algorithm to maximize $\Pr(X^P)$, the development is similar to that of Geman and Geman[6] with the important additions of a spectral dimension and the multigrid generalization. We exploit the Gibbs equivalence theorem[7] which recognizes the MRF as a random process with a Gibbs distribution. The advantage of adopting this approach is that we can now compute the **maximum a posteriori** (MAP) estimate that maximizes the probability $\Pr(X^P)$ by iteratively sampling from the local Gibbs distribution pertaining to each site $s \in S_P^{(\sigma_n)}$. Our algorithm, called the "partitioning algorithm," seeks to maximize the distribution:

$$\Pr(X_s^P = x_s^P \mid X_r^P = \hat{x}_r^P, r \neq s) = \Pr(X_s^P = x_s^P \mid X_r^P = \hat{x}_r^P, r \in G_s)$$

$$= \frac{1}{Z_s} e^{-\frac{1}{T} U_s(x^P, g)} \tag{3}$$

where $\{X^P = x^P\}$ means $\{X_{s_1}^P = x_{s_1}^P, \ldots, X_{s_M}^P = x_{s_M}^P\}$, $Z_s = \sum_{x^P \in \Gamma} e^{-\frac{1}{T} U_s(x^P, g)}$, and $U_s(x^P, g)$ is the energy interaction of site $s \in S_P^{(\sigma_n)}$ with the neighborhood G_s. The realizations $\hat{x}_r$ are estimates of x_r obtained by using the results of the previous iteration. Quite conveniently, the computations of Z_s are not necessary for purposes of obtaining a MAP estimate. We denote g to be the spatial and spectral realizations of G, which is in contrast to the spatial-only realizations of the simpler G considered by Geman and Geman.[6] Specifically, g_s are the spectral observations at site s, and g are the collections of these observations for all the sites.

The energy function $U_s(x^P, g)$ depends on difference measures, called "disparity measures," which indicate the dissimilarity between a site s and a neighbor r. Our specific energy function combines two disparity measures that measure spectral difference in two very different ways: the spectral angle and the Euclidean distance.[1] This energy function is designed to respond to disparities as measured by the spectral angle measure and the Euclidean distance. Our reason for using a combined measure is that we seek to increase the sensitivity of the algorithm without introducing undesirable clutter. Rather than using a single measure with a very sensitive threshold that would respond to one specific type of spectral difference, we use two measures that respond to different kinds of spectral difference with less sensitivity.

A simulated annealing technique is used to obtain the MAP estimate via the Metropolis algorithm,[1,2,5,6,7,9] with an eventual fast freeze at low temperature. The computations involve a temperature t_k at iteration k that defines an

annealing schedule. Furthermore, the multigrid implementation uses a grid-coarsening and cascading technique that is consistent with aspects of the Renormalization Group Algorithm (RGA) method developed for the restoration of degraded images.[8] Specifically, the fine-to-coarse sequence, described above, with partition lattices defined by (1) is reversed to a coarse-to-fine sequence: $S_P^{(\sigma_{N_{Seq}})} \rightarrow \dots \rightarrow S_P^{(\sigma_2)} \rightarrow S_P^{(1)}$. The results of stage "$n$" are used to initialize the algorithm at the next finer stage "$n-1$."

2.3 Initializing the Algorithm

The algorithm requires an initial estimate of the $\hat{x}_r^P$ realizations. These values can be initialized randomly, in which case during the first iteration the estimated realizations $\hat{x}_r^P$ are assigned class labels randomly. This is a worst-case scenario. Alternatively, we can consider some other means of initialization.

One alternative is to initialize the algorithm with the results of some prior supervised classification run. Consider (2) and recall our assumptions $X^\lambda = G$ and $P(G = g \mid X^P = x^P) = 1$ that simplifies (2) to $\Pr(X^P \mid X^\lambda) \propto \Pr(X^P)$. This was actually a gross simplification required to keep the algorithm analytically tractable and computationally practical so that we can incorporate spatial information into (2).

A common spectral-only Bayesian approach is to use (2) and assume that the site-specific distributions are independent so that (2) becomes $\Pr(X^P \mid X^\lambda) = \prod_s \Pr(X_s^\lambda \mid X_s^P)\Pr(X_s^P)$ and to further assume equal prior probabilities (all the $\Pr(X_s^P)$ are equal). Consequently, each site in an image can be classified independently of the other sites by selecting $\hat{x}_s^P$ which maximizes $\Pr(X_s^\lambda \mid X_s^P)$. Furthermore, these class-conditional distributions are often assumed to be multivariate normal, which leads to the classification rule:

$$\hat{x}_s^P = \arg\max N\left(\mu_{x_s^P}, \Sigma_{x_s^P}\right) \qquad (4)$$

where $\mu_{x_s^P}$ is the mean, and $\Sigma_{x_s^P}$ is the covariance, of the spectral vectors for the class label x_s^P, respectively. Eq. (4) is a well-known quadratic multivariate classifier used in many pattern recognition scenarios.[4,5] Although the multivariate normal assumption is often violated, it provides a far better approximation to the true distribution the class-conditional spectral data than the simplifying assumption $P(G = g \mid X^P = x^P) = 1$ made above.

The classification rule in (4) can be simplified further if we make additional assumptions on the class covariance matrices. Specifically, if we make the approximation that $\Sigma_{x_s^P} = I$, the Identity Matrix, for all $x_s^P \in \Gamma$, the classification rule simply reduces to the Euclidean distance classifier, which is a linear classifier. [4,5]

The partitioning algorithm has the advantage of incorporating spatial information and smoothing but the disadvantage of using a crude class-conditional distribution. The classifier provided in (4) uses a better approximation to the true class-conditional (spectral-only) distribution and can be shown to be the optimal classifyier when the underlying assumptions are met; however, it does not model any spatial interaction or provide any spatial smoothing that could be provided by a spectral/spatial prior distribution $\Pr(X^P)$.

Initializing the partitioning algorithm with the results of a classifier that computes (4) offers the potential for improved classification by providing initial estimates based on label estimates that under the proper conditions are spectrally optimal. Accordingly, we propose the following two-step procedure:

Step 1: Perform a classification of the scene using a supervised classifier of the form (4).

Step 2: Perform a spectral/spatial partitioning of the scene based on (3) initialized by the results in Step 1.

Implemented in this manner, the algorithm can also be viewed as a spectrally-optimal supervised classification algorithm that has been smoothed by a post-processing routine that imposes spectral/spatial constraints defined by the Gibbs prior probability distribution $\Pr\left(X^P\right)$.

3. DESCRIPTION OF EXPERIMENT

3.1. Data

A HYDICE scene is used for validation that was extracted from data collected over a geographic area near Fort Hood, Texas at a 10,000 feet altitude, on 16 May 1996. Figure 1 shows the scene, consisting of 300 samples by 600 lines of pixel vector data with a spatial resolution of approximately 1.5 meters. Because of the severe degradations that occur in atmospheric windows of the spectrum and sensor degradations occurring in certain bands, only 117 out of a possible 210 band were used to compute the disparities within the energy functions. These 117 bands were calibrated to reflectance by the Empirical Line Method based on the known spectra of the calibration panels that were located in a region outside this particular scene.

Figure 1. HYDICE Scene (Band 49)

Ground truth, describing certain ground features in the scene, was based on data collected during a site visit to the area during April 2000. In addition to the ground-feature data, spectral data of a diverse range of materials, measured by field spectrometers and collected independently of this experiment at other geographic locations, were used in prior work[1,2] to determine an appropriate quantization threshold for the spectral angle and Euclidean distance disparity metrics, which remained the same for this work. Ground-measured spectral data within the study site were not available.

3.2. Validation of the Algorithm

An experiment was conducted using computer code written in C++ that implements the algorithm described in Section 2. Two trials were conducted, investigating the merit of using the two-step supervised approach as compared to simply using the randomly initialized annealing approach.

In Trial A, the randomly initialized partitioning algorithm was run using the parameters shown in Table 1 for the various stages (grid sizes) of the multigrid process. We set $N_{Labels} = 10$. As indicated in Table 1, the trials proceeded in a multigrid sequence with grid resolutions $\sigma = 16, 8, 4, 2, 1, 1$ Simulated annealing was used to compute the MAP

estimates using an energy function comprised of the combined disparity measures that is identical to that used in our previous work. [1,2] A range $C = (0.05\text{-}3.0)$ was used, which is below the upper bound, and influences algorithm convergence as discussed in this previous work. The value of C was set to its highest value at the initial stage of the run (corresponding to high temperature) and it was set to lower values for the subsequent stages. In addition to the practical computational advantage, a lower C value allows the algorithm to "remember" the results at previous stages. The distances of the intermediate and far neighbors shown in Table 1 corresponds to fixed multiples of the grid resolution, 4σ and 8σ, respectively. The algorithm was initialized randomly, and then proceeded from a grid resolution of $\sigma=16$ to $\sigma=1$, using the extended neighborhood of 24 neighbors. A final pass was then made at $\sigma=1$ with a compact neighborhood comprised of 8 near neighbors.

Table 1. Control Parameters for Trial A. The sequence of numbers in the "Dist of neighbors" column corresponds to the distance from the center pixel (the "site") to the perimeter, for the surrounding near neighbors, the intermediate neighbors, and the far neighbors. The sequence of numbers in the "# of neighbors" column corresponds to the number of near neighbors, intermediate neighbors, and far neighbors in the neighborhood system.

σ	C	# iterations	Dist. of neighbors	# of neighbors
16	3.0	20,000	16,64,128	8, 8, 8
8	2.0	2,000	8,32,64	8, 8, 8
4	2.0	2,000	4,16,32	8, 8, 8
2	0.5	350	2,8,16	8, 8, 8
1	0.5	350	1,4, 8	8, 8,8
1	0.05	300	1	8

In Trial B, the two-step procedure was implemented. For Step 1, a supervised classification was performed using the Euclidean distance classifier. Recall that this classifier satisfies (4), given the assumption that $\Sigma_{x_s^P} = I$ for $x_s^P \in \Gamma$. We use this classifier rather than the quadratic classifier for two reasons: (1) We want to establish the degree of improvement over the randomly-initialized partitioning algorithm when using the simplest of assumptions on the class-conditional distributions, as well as to ascertain the capability of the partitioning algorithm to "clean up" the results of a rather simple classifier; (2) Using a quadratic classifier on the 117-band (very high-dimensional) HYDICE imagery requires the intermediate step of applying a dimension reduction technique (such as the Principal Component or the Minimal Noise Fraction transformation) to the data which implicitly adds another step that we have not discussed into the algorithm.

Using the information obvious in the scene supplemented by ground truth information 14 training classes were extracted from the scene. Some of these classes represent similar terrain features, such as Light Asphalt, Medium-toned Asphalt, and Dark Asphalt, which were subsequently combined after the classification run was finished. They were combined afterwards because combining such features into a single training class would either result in a multimodal class-conditional distribution or a unimodal class-conditional distribution of extremely high variance. So after the classification run was finished, the 14 classes were merged into 9 classes. The merged classes along with the corresponding original training classes are listed in Table 2. Further information about these classes is documented in ground-truth notes, but not discussed here.

For Step 2, the partitioning algorithm was tested two different ways. First, the algorithm was initialized by the results of Step 1 at a grid size of $\sigma=4$, and then set to run through the multigrid sequence down to $\sigma=1$ according to the parameters in Table 1. Second, the algorithm was simply initialized at the finest grid size $\sigma=1$ and then run at low temperature $(C=0.05)$ with the parameters shown in the bottom row of Table 1.

Table 2. Description of Classes.

Class Name	Description	# of samples
Asphalt	Light Asphalt, Med. Asphalt, Dark Asphalt	312
Concrete	Bright concrete	120
Roof L	Two light rooftops - Roof B1, RoofB2	162
Roof H	Rooftop of asphalt-singled houses	17
Roof B	Two brown rooftops - Roof J3, J2	80
Soil	Exposed soil	56
TreesDrk	Dark Green Trees	430
Grass	Two grass regions of differing vigor	727
GrassStr	Highly stress grass region	38

4. RESULTS

Figure 2 shows the resulting class map for the randomly initialized unsupervised partitioning algorithm in Trial A using the parameters in Table 1. The results are very similar to the results for N_{Labels} =6 obtained in our previous study. [1] Visually, the partitioning algorithm generates smooth partitions that accurately represent structured regions in the image. For most of the bigger regions, pixels with the same label represent the same type of phenomenon on the ground, globally across the image, and regions with different labels correspond to different phenomenon. However, a globally consistent labeling of the partitions was not completely achieved. In particular, labeling problems can be seen in some of the smaller regions. Although the contiguous pixels in smaller partitions represent the same phenomenon, locally, some small partitions, separated by a certain amount of distance and given the same names, actually represent different phenomenon. Conversely, some small partitions, separated by a certain amount of distance and given different names, actually represent similar phenomenon.

Figures 3 and 4 show the resulting class maps for Trial B. The class map in Figure 3 is the result of the spectral-only Euclidean classifier. Fourteen training classes were merged into the 9 thematic classes listed in Table 2. Visually, the map accurately portrays many of the structures in the scene and does a better job at extracting linear terrain features, such as roads, than the randomly initialized algorithm in Trial A; however, it does so with a severe amount of speckling. Figure 4 shows the results of the latter of the two implementations tested for Step 2 -- the implementation for which the algorithm was initialized at the finest grid size $\sigma=1$ and then run with the parameters shown in the bottom row of Table 1. Visually, this class map shows a significant improvement over those shown in Figures 2 and 3. The method retains the structures and the linear terrain features extracted by the Euclidean classifier while eliminating most of the speckling.

Regarding the other method tested for Step 2 (results not shown), this implementation tended to lose some of the linear features identified by the Euclidean classifier. Although the degradation was not severe, this approach is also computationally more expensive. So, whereas a multigrid implementation is the preferred approach when the algorithm is initialized randomly, the latter (single pass) implementation is preferred for the two-step method.

Tables 3 and 4 provide supporting quantitative results regarding the labeling in the Trial B class maps. Tables 3a and 3b provide auto-classification results, listing the percentage of sites in each of the training regions that have been assigned to a particular training class, for Step 1 and Step 2, respectively. Tables 4a and 4b provide classification results for test data extracted from the scene based on ground-truth, listing the percentage of sites in each of the test regions that have been assigned to a particular training class.

Figure 2. Trial A -- Randomly-Initialized Multigrid Partitioning Algorithm.

Figure 3. Trial B (Step 1) – Spectral-Only Supervised Classification.

Figure 4. Trial B (Step 2) – Partitioning Algorithm Initialized by Supervised Classification.

Table 3. Quantitative Results for Test Data in Trial B. The columns list the percentage of sites in each of the training regions that have been assigned to a particular training class (i.e. Training data are tested against itself). Percentages highlighted in bold (and expected to be on the diagonal) are considered to be correct assignments.

Table 3a. Results for the Euclidian Distance Classifier.

	Asphalt	Concrete	Roof L	Roof H	Roof B	Soil	Trees	Grass	GrassStr	#pts
Asphalt	**99.7%**	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%	0.0%	0.0%	312
Concrete	6.7%	**54.2%**	31.7%	0.0%	0.0%	7.5%	0.0%	0.0%	0.0%	120
Roof L	1.2%	4.9%	**92.6%**	1.2%	0.0%	0.0%	0.0%	0.0%	0.0%	162
Roof H	0.0%	0.0%	5.9%	**94.1%**	0.0%	0.0%	0.0%	0.0%	0.0%	17
Roof B	6.3%	0.0%	0.0%	0.0%	**93.8%**	0.0%	0.0%	0.0%	0.0%	80
Soil	0.0%	1.8%	0.0%	0.0%	0.0%	**98.2%**	0.0%	0.0%	0.0%	56
TreesDrk	3.5%	0.0%	0.0%	0.0%	5.8%	0.0%	**63.5%**	27.2%	0.0%	430
Grass	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%	**95.9%**	0.4%	727
GrassStr	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**100.0%**	38

Table 3b. Results for the Partitioning Algorithm Initialized by the Euclidian Distance Classifier.

	Asphalt	Concrete	Roof L	Roof H	Roof B	Soil	Trees	Grass	GrassStr	#pts
Asphalt	**100.0%**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	312
Concrete	0.8%	**86.7%**	12.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	120
Roof L	0.0%	0.0%	**100.0%**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	162
Roof H	0.0%	0.0%	0.0%	**100.0%**	0.0%	0.0%	0.0%	0.0%	0.0%	17
Roof B	0.0%	0.0%	0.0%	0.0%	**100.0%**	0.0%	0.0%	0.0%	0.0%	80
Soil	0.0%	0.0%	0.0%	0.0%	0.0%	**100.0%**	0.0%	0.0%	0.0%	56
TreesDrk	2.1%	0.0%	0.0%	0.0%	0.0%	0.0%	**82.8%**	15.1%	0.0%	430
Grass	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%	**99.3%**	0.0%	727
GrassStr	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**100.0%**	38

These tables demonstrate a clear benefit to using the algorithm. Table 3 shows that when the algorithm is applied to its own training data, the percentage of correct labeling increases for eight out of the nine classes. For the ninth class, the classification accuracy was already 100% so there was no possibility of improvement. Table 4 shows that when the algorithm is applied to test data from regions outside of the training data there is consistent and significant improvement in accuracy. Out of the 15 test regions, there was improvement in nine of the regions, there was no improvement for three of the regions, there was no chance of improvement for two of the region (100% accuracy), and a decrease in one of the regions. Notice that the degree of improvement was sometimes substantial. For example, the accuracy of the test region for trees improved from 66.9% to 87.9%.

In the case where there was a decrease in accuracy, the partitioning algorithm was still behaving in an appropriate manner. The spectral signatures of concrete and soil are quite similar (concrete being composed of soil/silicate materials) and consequently simple classifiers often confuse these two classes. In fairness to the algorithm, in this case, the Euclidean classifier incorrectly labeled 66.7% of the soil region as concrete and another 10.7% as asphalt, leaving only 20.1% of the sites correctly labeled for the initialization step. Understandably, the partitioning algorithm then proceeded to expand the incorrect concrete labeling to 91.8%.

Table 4. Quantitative Results for Test Data in Trial B. The columns list the percentage of sites in each of the test regions that have been assigned to a particular training class. Percentages highlighted in bold are considered to be correct assignments.

Table 4a. Results for the Euclidian Distance Classifier.

	Asphalt	Concrete	Roof L	Roof H	Roof B	Soil	Trees	Grass	GrassStr	#pts
AsphaltLgt_T	**80.4%**	0.0%	0.0%	0.0%	18.5%	0.0%	0.0%	1.1%	0.0%	92
AsphaltMed_T	**86.6%**	0.0%	0.0%	0.0%	13.4%	0.0%	0.0%	0.0%	0.0%	97
AsphaltDrk_T	**100.0%**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	437
RoofB1_T	0.0%	0.0%	**100.0%**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	173
Concrete_T	0.0%	**87.5%**	12.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	24
RoofGravel	11.8%	**82.4%**	5.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	17
RoofB2_T	6.3%	17.2%	**76.6%**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	64
TennisCourt	2.3%	0.0%	0.0%	97.7%	0.0%	0.0%	0.0%	0.0%	0.0%	43
RoofHouse_T	0.0%	0.0%	24.4%	**66.7%**	8.9%	0.0%	0.0%	0.0%	0.0%	45
RoofJ3_T	5.1%	0.0%	0.0%	0.0%	**94.9%**	0.0%	0.0%	0.0%	0.0%	39
RoofJ2_T	7.1%	0.0%	0.0%	0.0%	**92.9%**	0.0%	0.0%	0.0%	0.0%	42
SoilLgt_T	0.0%	11.6%	0.0%	0.0%	0.0%	**88.4%**	0.0%	0.0%	0.0%	138
SoilTan_T	10.7%	66.7%	0.6%	0.0%	0.0%	**20.1%**	0.0%	0.6%	1.3%	159
GrassWild_T	0.2%	0.0%	0.0%	0.0%	0.7%	0.0%	18.9%	**74.7%**	5.5%	1700
TreesDrk_T	5.5%	0.0%	0.0%	0.0%	11.6%	0.0%	**66.9%**	16.0%	0.0%	1591
GrassStr_T	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**100.0%**	35

Table 4b. Results for the Partitioning Algorithm Initialized by the Euclidian Distance Classifier.

	Asphalt	Concrete	Roof L	Roof H	Roof B	Soil	Trees	Grass	GrassStr	#pts
AsphaltLgt_T	**84.8%**	0.0%	0.0%	0.0%	14.1%	0.0%	0.0%	1.1%	0.0%	92
AsphaltMed_T	**96.9%**	0.0%	0.0%	0.0%	3.1%	0.0%	0.0%	0.0%	0.0%	97
AsphaltDrk_T	**100.0%**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	437
RoofB1_T	0.0%	0.0%	**100.0%**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	173
Concrete_T	8.3%	**87.5%**	4.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	24
RoofGravel	0.0%	**100.0%**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	17
RoofB2_T	0.0%	0.0%	**100.0%**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	64
TennisCourt	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	43
RoofHouse_T	0.0%	0.0%	22.2%	**77.8%**	0.0%	0.0%	0.0%	0.0%	0.0%	45
RoofJ3_T	0.0%	0.0%	0.0%	0.0%	**100.0%**	0.0%	0.0%	0.0%	0.0%	39
RoofJ2_T	0.0%	0.0%	0.0%	0.0%	**100.0%**	0.0%	0.0%	0.0%	0.0%	42
SoilLgt_T	0.0%	11.6%	0.0%	0.0%	0.0%	**88.4%**	0.0%	0.0%	0.0%	138
SoilTan_T	3.8%	91.8%	0.0%	0.0%	0.0%	**1.9%**	0.0%	0.6%	1.9%	159
GrassWild_T	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	20.2%	**74.8%**	5.0%	1700
TreesDrk_T	3.0%	0.0%	0.0%	0.0%	3.1%	0.0%	**87.9%**	6.0%	0.0%	1591
GrassStr_T	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**100.0%**	35

The results show that spatially smooth labeling can be achieved without decreasing the accuracy of classification. Tables 3 and 4 show an overall increase, not decrease, in classification accuracy when going from the spectral-only to the spectral/spatial process. Furthermore, Figure 4 shows that much of the speckling and edge artifacts in a scene labeling can be removed without removing spectrally significant individual pixels. For example, observing the parking lot adjacent to the department store (upper-left corner) in Figure 4, the single vehicles are not removed, whereas the edge-artifacts of the roads and the clutter in the forest and grass regions throughout the scene are mostly removed.

5. CONCLUSIONS

A Bayesian framework is used to develop a 2-step supervised classification algorithm that is capable of performing high-quality spatially smooth labeling of hyperspectral imagery. To demonstrate the concept, a linear classifier was used to initialize a Gibbs-based partitioning algorithm resulting in significantly improved label accuracy and smoothness as compared to the linear classifier itself. In addition, the global labeling accuracy was increased as compared to the stand-alone randomly initialized partitioning algorithm. Based on optimality arguments, the current results will likely improve further, if the linear classifier is replaced with a quadratic classifier.

In order to achieve a reasonable global labeling accuracy, the randomly initialized (unsupervised) partitioning algorithm is best implemented as a multigrid process using an extended neighborhood system. However, initializing the partitioning algorithm with even the simplest of spectral-only supervised classifiers not only improves the global accuracy, but it also reduces the computation by (1) eliminating the need for a multigrid process and (2) allowing the annealing to start at a cooler temperature.

The experiment shows that spatially smooth labeling does not have to occur at the cost of classification accuracy, which has been the case with a number of post-processing methods that simply apply spatial constraints to a spectral-based classification.
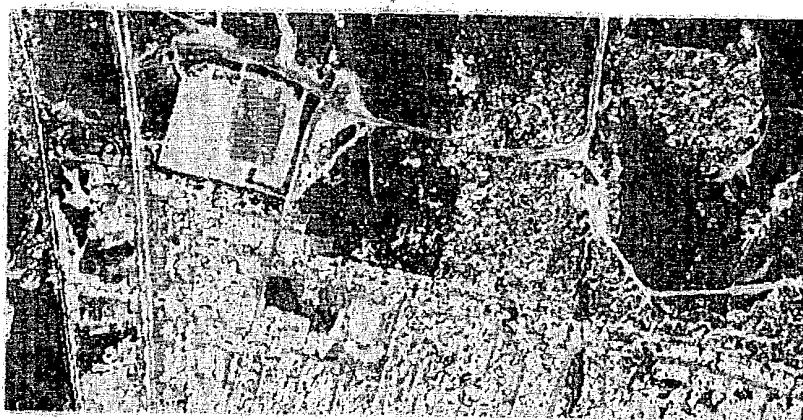
ACKNOWLEDGEMENTS

REFERENCES

[1] R. Rand and D. Keenan, "A Gibbs-based unsupervised segmentation approach to partitioning hyperspectral imagery for terrain applications," Proceedings of the SPIE Aerosense, Orlando, FL., April 2001.

[2] R. Rand and D. Keenan, "A Spectral Mixture Process Conditioned by Gibbs-based Partitioning," *IEEE Transactions on Geoscience and Remote Sensing: Special Issue on the Analysis of Hyperspectral Image Data*, Vol. 39, No.7, July 2001.

[3] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 1973.

[4] C. Therrien, *Decision Esimation and Classification – An Introduction to Pattern Recognition and Related Topics*, John Wiley and Sons, New York, 1989.

[5] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods – A Mathematical Introduction*, Applications of Mathematics, Springer-Verlag, New York, 1995.

[6] D. Geman, S. Geman, C. Graffigne., and P. Dong, "Boundary detection by constrained optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 7, July 1990.

[7] D. Geman and S. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, November 1984.

[8] B. Gidas, "A Renormalization Group Approach to Image Processing Problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, No. 2, February 1989.

[9] E. Aarts and J. Korst, *Simulated Annealing and Bolzmann Machines - A Stochastic Approach to Combinatorial Optimization and Neural Computing*, Interscience Series in Discrete Mathematics and Optimization, John Wiley and Sons, 1989, Reprint 1990.
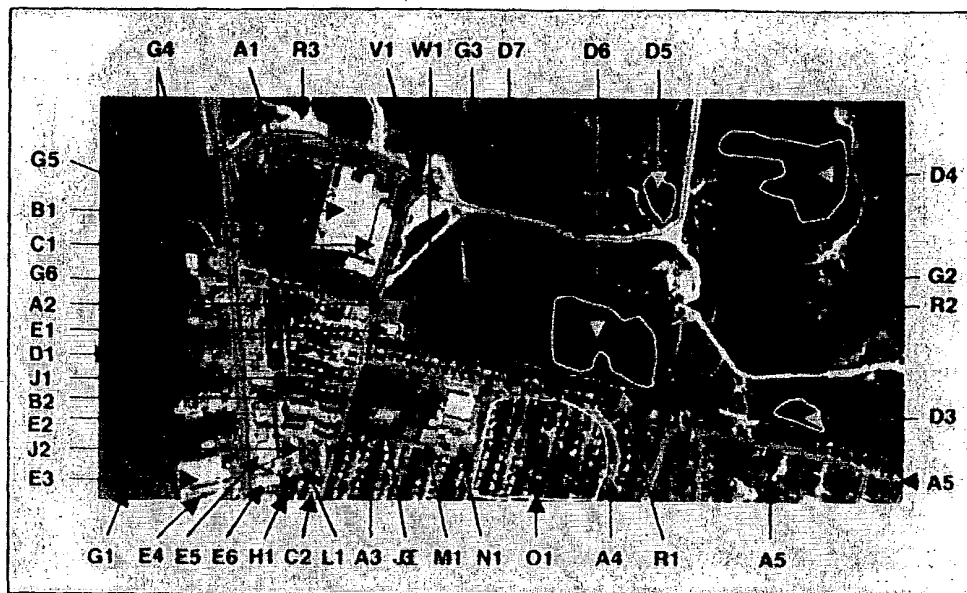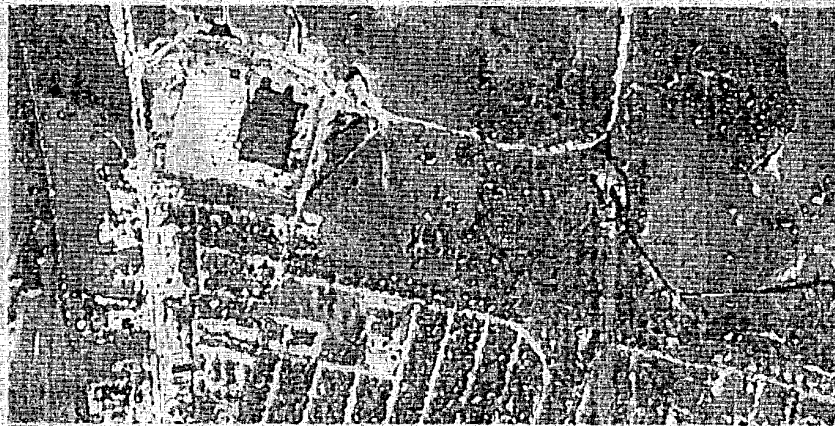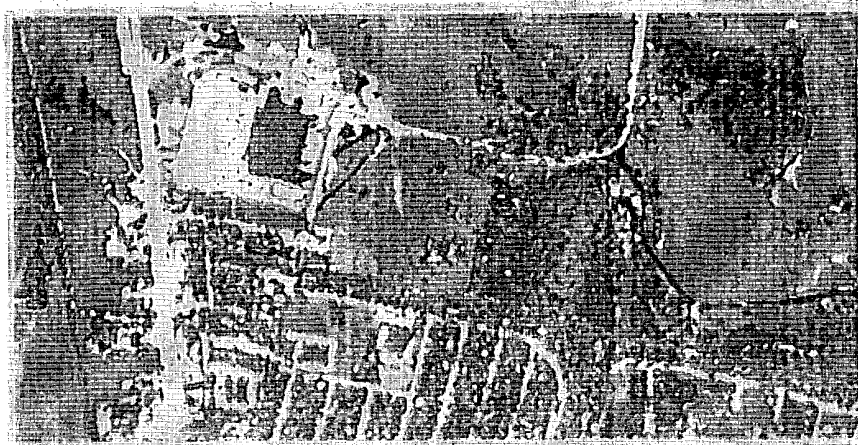
Fig. 1

Fig. 2

Fig. 3

Fig 4

Fig. 5

Fig 6

Fig. 7